DOCUMENT RESUME

ED 178 348 SE 029 285

AUTHOR	Banaugh, R. P.
TITLE	An Introduction to Computer Assisted Analysis in the Biological Sciences.
INSTITUTION	Montana Univ., Missoula. Dept. of Computer Science.
SPONS AGENCY	National Science Foundation, Washington, D.C.
PUB DATE	30 Sep 76
GRANT	NSF-GZ-3469
NOTE	619p.; Page 1.9 missing from document; Contains occasional light and broken type
EDRS PRICE	MF03/PC25 Plus Postage.
DESCRIPTORS	*Biological Sciences; *College Science; Componential Analysis; Computer Assisted Instruction; *Computer Oriented Programs; Computer Programs; Genetics; Growth Patterns; Higher Education; *Models; Population Growth; Problems; Textbooks

ABSTRACT

This set of notes is designed to introduce the student to the development and use of computer-based models, and to analyze quantitative phenomena in the life sciences. Only BASIC programming language is used. The ten chapter titles are: The Growth of a Single Species; The Association of Two Species; Parameter Determination; Automated Parameter Determination; Multivariable Search Methods; Life Tables; Applications to Genetics; Random Processes; Compartmental Analysis; and Simulating Tree Growth with a Computer. Most chapters include problem and reference sections. (MK)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

REC'D SEDR/MCD NOV 8 1976

File →

Archives of
HIGHER EDUCATION IN SCIENCE
ATP, SDIP, [illegible]

Property of
THE NATIONAL SCIENCE FOUNDATION

GZ-3469
(COMP. SCIENCE)

ED178348

An Introduction to Computer Assisted Analysis in the Biological Sciences

by

R. P. Banaugh

Department of Computer Science
University of Montana
Missoula, Montana

Prepared for the National Science Foundation
Washington, D. C. 20050

Contract No. GZ-3469

September 30, 1976

U S DEPARTMENT OF HEALTH
EDUCATION & WELFARE
NATIONAL INSTITUTE OF
EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY
Mary L. Charles
NSF
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

SE 029 285

TABLE OF CONTENTS

PREFACE i

CHAPTER I. THE GROWTH OF A SINGLE SPECIES

Introduction 1.1
The Malthus Problem 1.3
Effect of a Fixed or Constant Resource 1.9
Contamination of the Environment 1.13
Retarded Time Effects 1.16
Effect of Mating Possibility 1.22
Summary 1.25
Appendix 1.26
Remarks Concerning the Problems at the End of the Chapters 1.36
Problems 1.37
References 1.41

CHAPTER II. THE ASSOCIATION OF TWO SPECIES

Independent Growth 2.1
Effect of Fixed or Finite Resources 2.2
One Species Feeds Upon the Other 2.10
A Leslie Type Model 2.17a
Effect of Emigration 2.18
Environmental Toxicity 2.20
Other Models 2.24
Running of Programs 2.32
Closing Comments 2.37
Problems 2.39
References 2.42

CHAPTER III. PARAMETER DETERMINATION

Introduction 3.1
An Example 3.2
Comparison of Two Curves 3.3
Comparison With Experiment 3.8
A Suggested Method 3.8
The Program Modifications 3.11
Problems 3.21

CHAPTER IV. AUTOMATED PARAMETER DETERMINATION

Introduction 4.1
Minimization 4.1
An Automated Iterative Procedure on Algorithm . . . 4.7
A Minimization Algorithm 4.8
Starting Values 4.14
Program Results 4.17
Interpolation 4.22
Problems 4.28
Reference 4.30

CHAPTER V. MULTIVARIABLE SEARCH METHODS

Introduction 5.1
An Example 5.1
Starting Values and Results 5.2
General Comments 5.9
A Minimization Algorithm 5.11
Modifications 5.25
Starting Values 5.26
Numerical Results 5.27
Problems 5.32
References 5.35

CHAPTER VI. LIFE TABLES

Introduction 6.1
Development of the Model 6.2
An Example 6.10
A Pictorial Representation 6.26
Harvest 6.33
Discussion 6.39
Population Effects (Effect of a Finite Resource . . 6.41
Effect of Mating Possibility 6.45
Problems 6.48
References 6.53

CHAPTER VII. APPLICATIONS TO GENETICS

Some Preliminaries 7.1
Mendel's Experiments 7.2
The Prediction of Heredity Characteristics 7.9
The Development of the Computer Program 7.14
Discussion of Some Results 7.19
Simple Extensions 7.24
The Hardy-Weinberg Principle 7.38
Gene Dispersal 7.51
Related Projects 7.55

Small Population 7.56
Program Description 7.66
Miscellaneous Comments 7.83
Program Results 7.84
Another Program Alteration 7.88
Other Extensions 7.107
Comments on the Chapter Problems 7.112
Problems 7.113
References 7.117

CHAPTER VIII. RANDOM PROCESSES

Introduction 8.1
A Very Simple Problem 8.2
A Second Simple Problem 8.11
A Third Problem 8.15
A Fourth Problem 8.18
A Fifth Problem 8.21
A Sixth Problem 8.31
Seventh Problem 8.36
The Eighth Problem 8.38
Ninth Problem 8.41
Discussion 8.48
Appendix A, The Relation to Classical Probability Theory 8.50
Appendix B 8.75
Problems 8.85
References 8.91

CHAPTER IX. COMPARTMENTAL ANALYSIS

Introduction 9.1
Transfer of the Fluid from the Muscle Tissue 9.4
Example #2 . . . . 9.19
Example #3 . . . . 9.31
The Tracer Method 9.36
Determination of Transfer Coefficients 9.44
An Illustration of a Non-Linear Model 9.73
Food Chain Kinetics 9.75
A Nuclear Fallout Problem 9.76
A Second Example in Food Chain Kinetics 9.82
The Movement of a Pesticide in a Food Chain 9.97
Relation of this Formulation to the Differential Equation Formulation 9.108
Problems 9.119
References 9.122

CHAPTER X. SIMULATING TREE GROWTH WITH A COMPUTER

by W. R. Pierce

Abstract

Material 10.1
Literature Cited 10.30

Preface to the Student

The purpose of this set of notes is to introduce the student to the development and use of computer based models, and to analyze quantitative phenomena in the life sciences. This will be accomplished by using only the BASIC programming language.

Prior to the advent of the modern high-speed digital computer, the prediction and quantitative analysis of natural or life science phenomena was a most difficult endeavor. This was due to the fact that effective mathematical descriptions of biological phenomena frequently resulted in equations which were extremely difficult to solve. The development of the digital computer and the use of numerical methods of solution has rekindled interest in quantitative methods. This, in turn, has inspired a surge of new courses with such titles as, "Mathematics for the Biological Sciences", "Mathematics for the Life Sciences", etc. The goal of these courses is to provide the students with mathematical techniques, which hopefully, will assist biology students in quantitative or mathematical methods of analysis. It is the purpose of these notes to present an alternative and complimentary technique to assist in the theoretical analysis of biological phenomena.

The general procedure in pursuing a theoretical investigation is to carefully isolate and clearly identify the problem and to then formulate specific hypotheses concerning the essential phenomena. Utilizing the mathematical equivalents of these hypotheses, a set of equations is derived whose solutions should yield the desired information. In all but the most trivial of cases, the equations are usually intractable and must be solved with the aid of a digital computer. It is then necessary to use numerical methods, e.g. finite difference methods, to transform the mathematical equations into a form which permits the solution to be effected by a computer. Finally, a computer program is developed and the program run on the computer to obtain the results.

In the course of following such a procedure, it is almost always the case that the computer program must be altered as initial conjectures or hypotheses are altered or as the availability of experimental data changes. Such alterations usually begin by rewriting the mathematical expressions or equations to accommodate the altered hypothesis.

Next, the numerical and then the computerized form of these equations is corrected and finally the program is altered to effect the changes in the hypotheses. As the program continues to be developed and evaluated, the investigator more and more proceeds to directly alter the program, thus omitting the intermediate step of changing the mathematical description. The individual lines and terms of the program become, in a sense, a description or expression of the behavior of the phenomena and this description is just as meaningful as the original mathematical description. This then suggests that it may be possible to proceed directly to a description of the phenomena in terms of the programming language and to omit the intermediate mathematical stage. It is just this procedure that is followed in this text.

In this work, a deliberate attempt is made to minimize a knowledge of formal mathematics in an effort to see just "how far" we can go in presenting to the student, in a pedagogically sound manner, rational methods of quantitatively analyzing biological phenomena. This is not an effort to downgrade or minimize the importance of mathematics (this point is addressed again in the introduction and other places in the text), rather it is just what it sets out to be; i.e. an attempt to develop quantitative methods of analyzing phenomena in the natural sciences using only a simple programming language. The language BASIC has been chosen because of its wide acceptance in schools, its interactive capability, and its availability on both minicomputers and larger computers. The more perceptive student and the student who further investigates these methods may choose to use some other programming languages such as FORTRAN, PL1, APL, etc.

Since it is the intent of the text to train the student in the formulation, development and analysis of computer programs describing biological phenomena, stress will be laid upon:

(a) The establishment of a clear and unambiguous statement of the problem,

(b) The careful and complete specification of the necessary or relevant hypotheses,

(c) An analysis of the computer output or results and the role that such analysis plays in suggesting new alterations to the program. The development of the computer program and the analysis of the results is often times a very fruitful source of new ideas to assist in the obtaining of a more realistic explanation of the problem.

Another important reason for proceeding in this manner is to take advantage of the well-known pedagogical value of a computer. The very act of defining the problem, formulating a method of solution and programming and debugging the program has been found to be an excellent technique for furthering the student's understanding of the phenomena under investigation.

The increasing use of the computer in all disciplines has resulted in the development of programs which are very large and sophisticated and are applicable to many problems other than those for which the programs were originally developed. In addition, there are problems which are of frequent occurrence and special, very efficient computer programs have been constructed to assist in the solution of these problems. Such programs are called, "Canned Programs" and are usually available without cost or for a very small nominal fee. It is the contention of your author that the effective use of these programs is most readily obtained if the user has had experience in developing and writing such programs; if only on a limited scale. A user of such programs should have a skeptical attitude and the experience to appreciate and recognize all of the limitations, as well as the capabilities, of the program. Consequently, a further reason for the student learning how to develop computer models is to enable him to better use canned programs. Finally, the feasibility of proceeding in this manner is in keeping with the thesis that if a hypotheses can be unambiguously stated and rationally defended then it is possible to program it and to then examine the consequences with the aid of the digital computer.

The presentation in the text will be almost entirely by example. In the earlier chapters, biological phenomena which readily and easily lend themselves to such a technique of analysis, will be investigated and later chapters will discuss more difficult and complicated phenomena. The style of the presentation is informal because this is an introductory course whose purpose is to give the student facility and appreciation of computer assisted analysis. Definitions of terms and notations are frequently repeated in discussions to better fix such definitions in the student's mind. Such repetition also enables the student to more easily follow the discussion because it minimizes interruptions caused by the necessity to refer back to the original mention of the term or notation. In addition, because the aim of the work is to enable the student to obtain an understanding of computer based quantitative analysis, most of the computer programs developed or presented are not optimal. They usually can be shortened, made more efficient or dressed up in some manner.

The development of problems for such a text is quite difficult and consequently only a limited number appear. However, frequent suggestions are given in the body of the text for altering the hypotheses and hence the computer programs and their analysis. An evaluation of a student's ability to effect such suggestions should serve as a partial gude to the measure of the student's grasp of the subject matter. Also the assignment of projects, usually student originated, should further assist the student's understanding and capability.

The primary difficulty encountered by students in attaining familiarity with quantitative methods in the biological sciences is quite analogous to the difficulty students encounter in attaining a corresponding understanding of quantitative methods in the physical sciences. For these latter students, this is the difficulty of correctly stating or applying the mathematical statements of the underlying physical principles. Thus, it is not surprising that the student of quantitative methods in the biological sciences encounters a like hardship. The student will discover that the primary difficulty will be the expressing in the BASIC programming language the quantitative consequences of his assumptions concerning the biological phenomena. This difficulty will reflect itself in the delineation of the problem as well as in the design and implementation of the program. Thus, in this work we have tried to emphasize problem definition and program formulation.

It may be of interest to remark that the modern high-speed digital computer is expected to accelerate the convergence of the efforts of the biological and the physical sciences. Physics has sometimes been defined as the science which seeks to formulate underlying quantitative principles about our universe. As a consequence, the development of techniques for optimally applying these principles has been called engineering. In the past, the necessity and desire for mankind to improve his lot here on earth resulted in great emphasis and interest being placed on the discovery and utilization of these physical principles.

The dramatic awareness of mankind of the finiteness of our local universe and specifically of all that this awareness implies, has shifted the emphasis from mankind trying to unrestrictedly improve its lot, to mankind trying to assure the continuance of what it has achieved. The recognized urgency of this shift of goals, coupled with the fortuitous development of the modern high-speed digital computer, has given a very strong impetus to the development of analogous underlying quantitative principles in the biological sciences. When restricted to natural resources, such study has sometimes been called ecology. The application of these principles has been termed ecological engineering.

It is assumed that the student is familiar with the BASIC programming language or that a short period of time is set aside at the beginning of the course to present the rudiments of BASIC. The ready availability of computer facilities providing the BASIC language is a necessity; most preferably in the time share mode.

CHAPTER I

THE GROWTH OF A SINGLE SPECIES

Introduction

As a first example of a model or a simulation of a phenomenon in the life sciences, we consider the problem of describing the growth of a population. This is an old problem and one of the first biological phenomena to be mathematically treated. V. Volterra, an Italian mathematician, was one of the earliest investigators of such problems. His work began about 1926 and it is his work we shall follow.

In developing this and other models, various suppositions and assumptions will be made. It is important that the student understand and appreciate the implications and limitations of each of the assumptions since any alterations of these basic hypotheses may significantly alter the predictions obtained from the model. As digital computers had yet to be developed when Volterra did his pioneering work, his models were expressed in the language of mathematics. In particular, he used the differential and integral calculus and the subjects of differential and integral equations quite extensively. Because this is a course in computer applications in the life sciences, we will use the programming language BASIC to express our model in contrast to the language of mathematics. There are advantages in the use of either language. One significant advantage of the mathematical formulation is the opportunity it affords to bring the vast framework of mathematical knowledge to bear upon the problem. In addition, when truly complicated models are proposed, the mathematical formulation can sometimes be used to discover biologically significant relations without the expense of considerable computer time. Furthermore, mathematics enables the investigator to discern common structures of different models and hence serves as a unifying tool for science. On the other hand, the programming language BASIC is extremely easy to learn and to use and is readily understood by a far larger audience. In addition, since almost all sufficiently complicated mathematically formulated problems have to be solved with the aid of a computer anyway, there may be a significant shortening of time required to develop the computer program by a directly formulating problem in terms of the programming language. In addition, the use of the infinitesmal calculus requires the assumption that the variables change in a smooth or continuous manner. In contrast, a programming language naturally permits finite or step-wise changes in the variables. Consequently, for discrete phenomena such as population growth, computer simulation may require less restrictive assumptions and permit a more realistic formulation than could be obtained by using mathematics. As stated previously, it is our intent to proceed as far as possible using just a simple computer language. For the student who wishes to learn more sophisticated modeling techniques, certainly a strong mathematical background may be necessary and it is our hope that this course will encourage the student to obtain such a background.

In building our model for the simulation of the growth of a single population we will begin by making some very, very simple assumptions. For this reason, the model will be quite oversimplified; however, it will permit a ready and easy programming and running on the computer. Thus, we will be able to quickly make several computer runs. The examination of the results together with our knowledge of "what we left out" of the original hypothesis will then suggest alterations and/or additions to the original model. The process will then be repeated again and as often as is necessary in order to arrive at what we hope or believe is a realistic explanation or model of the growth of a population. As the model progresses in complexity and growth, it is important that the student "sees" or is aware of the contribution to the building of the model that an analysis of results on simpler models provides. In other words, the computer model itself becomes a powerful tool in suggesting alterations and additions to the model. This effect is not so noticeable in the early or primitive stage of the model building. However, as the complexity and completeness of the description of the population growth increases, the analysis of the computer results may be the only means of suggesting new alterations or additions.

The fundamental idea we shall use in our model-making or simulation is the following:

The next state of the variable equals the old or preceding state of the variable, plus the change in the variable.

In the language of BASIC this can be stated in the following way:

```
10  LET V(I+1) = V(I) + C(I)
```

where V(I+1) and V(I) are the new and old states of the variable respectively, and C(I) is the change in the variable. This law and variations of it will form the basis of most of the development in this work. The importance and utility of the law cannot be overstressed and for this reason we call it the Law of Change.

By identifying the variable V(I) with the population at the i^{th} time period we see that the above law states, "the population at the beginning of the next time period is equal to the population at the beginning of the present time period plus the change in population in this time period."

Hence, to apply this law to our population growth model, we need to specify two quantities:

i) The population at the initial time.
ii) The amount of change in any time period.

The Malthus Problem

The complexity of the model will be determined by our hypothesis concerning the second quantity. It is, of course, readily apparent that such factors as environment, season, time of day, prey and/or predator specie, age, sex, etc. will all influence our choice or specification of the rule of change. Furthermore, the length of the time period will directly affect the magnitude of the change, C(I). For reasons given below we wish to make our initial model as simple as possible and yet permit it to have some creditability. Thus, we will adopt the convention of equal time periods. This is an assumption and, as far as computational or programming difficulty is concerned, it is not necessary. The constancy of the time period is adopted purely for convenience in getting started on our model building. In the following, the student will frequently find it most helpful to think of the time period as the generation period; i.e. the period of time between generations. In order to determine a simple, yet not trivial expression for the change in the population in a time period, we examine in a bit more detail what we are trying to do.

We would like to end up with a model that is realistic and also quite inclusive, yet we suspect that such a model will be very complicated and not easy to come by. Realizing that complicated models are composed of many parts and further realizing that a sensible way of building a complicated or realistic model is to start with a simple yet thoroughly understood model and to then refine it, we begin by trying to select assumptions that will, we hope, lead to such a simple model. In addition, to the advantages mentioned above, a simple "first" model has other advantages, e.g. it is easy to check results obtained from the simple model. Thus, our intuition is either confirmed or rejected early in the process. With these ground rules in mind we now proceed to describe Volterra's initial assumptions concerning the growth of a population.

It is certainly true that the change in the population in any given period of time is equal to the number of births minus the number of deaths. Thus, we can write

$$C(I) = \begin{pmatrix}\text{No. of births}\\ \text{in the period}\end{pmatrix} - \begin{pmatrix}\text{No. of deaths}\\ \text{in the period}\end{pmatrix}$$

In order to arrive at expressions for the number of births and the number of deaths in a period, certain assumptions must be made. It is quite evident that a host of elements affect each of these expressions. Some of them are: weather, polution, health, availability of food and water, stress, etc. It is perhaps more evident that a rational inclusion of the combined effect of all these elements is impossible. Consequently, it is necessary to make some very restrictive assumptions.

Volterra supposed that the plant or animal population, that we are attempting to describe, lived in isolation and in a constant environment. The term "constant environment" refers to the fact that during the entire time of the evolution of the population, all environmental factors such as food supply, living space, birth and death rates, etc. remain constant for each member of the population. The student should recognize the "real world" implications of such assumptions. For example, the assumption that the food supply remains constant for each individual in the population implies that if the population increases the total food supply must also increase in order that each individual may have the same amount of food. It may be possible to insure a constant food supply for a finite population, such as a cattle herd whose population is controlled; however, it is certainly not possible to provide a constant amount of food to each individual in an ever-increasing population. Thus, we expect that the results of our model might apply only to the early time growth of a population. In this case, we say that our model "breaks down", and hence is valid, only for early time histories. Volterra furthermore assumed that there was no intervention by any other species of plants or animals. These are assumptions, perhaps realizable for a short time in a laboratory, and are certainly ideal and not realistic. Nevertheless, they are not too far fetched and have the property of being quite precise. In line with these strict assumptions, Volterra further assumed that the number of births as well as the number of deaths in a unit of time was proportional to the total number of individuals existing during the unit of time. This is not an unreasonable assumption since for a given period of time, the greater the population the greater the number of births and deaths and the smaller the population the smaller the number of births and deaths. To express these notions in the BASIC language, we introduce the following notation for our variables:

Let P(I) denote the population at the beginning of the Ith time period. P(I) is assumed to be constant in this time period.

Let B be the constant of proportionality (or the fraction) relating the population to the number of individuals born in a time period. Thus, B*P(I) is the total number of individuals born in the Ith time period. B is called the coefficient of natality.

Let M be the constant of proportionality for the number of individuals who die in a time period. Thus, M*P(I) is the total number of individuals who die in the Ith time period. M is called the coefficient of mortality.

C(I), the total change in the population in the Ith time period, is then B*P(I)-M*P(I) and hence an application of the law of change gives

$$P(I+1) = P(I) + (B-M)*P(I) \quad (1)$$

The term B-M is called the <u>growth coefficient</u> and is denoted by G.

If we know the population at some initial time and are able to specify the constants B and M, we can write a simple program to evaluate our model. One such program is:

```
 1  REM POPULATION GROWTH MODEL, SIMPLE MODEL
10  DIMENSION P(60)
20  INPUT B, M, P(0)
25  LET G = B-M
30  FOR I = 0 TO 50
40  LET P(I+1) = P(I)+G*P(I)
50  NEXT I
60  FOR I = 0 TO 50
70  PRINT I, P(I)
80  NEXT I
90  END
```

The student is urged to run this program for various values of B, M, and P(0), the initial or starting population. (See figures 1.1a and 1.1b for some sample results).

The set of numbers, I and P(I), for each run is called the <u>population curve</u>. Upon examination of the results, the student will notice that if B>M, P(I) increases, i.e. P(I+1)>P(I) for all I. Moreover, the student will also notice that the increase in population per time period increases, i.e. (P(I+2)-P(I+1))>(P(I+1)-P(I)). Such growth is commonly called <u>exponential growth</u>. If B=M, there is no change in the population and if B<M, then the population approaches zero. These relations suggest that it is the quantity (B-M) that is important. These results all agree with our intuition since the statement B>M means there are more births than deaths in a time period and hence over a long enough period of time the population should increase without bound. The condition B<M means there are more deaths than births in a time period and we would thus expect that the population should eventually die out. Consequently, our simple model is not

SIMPLE POPULATION GROWTH MODELS

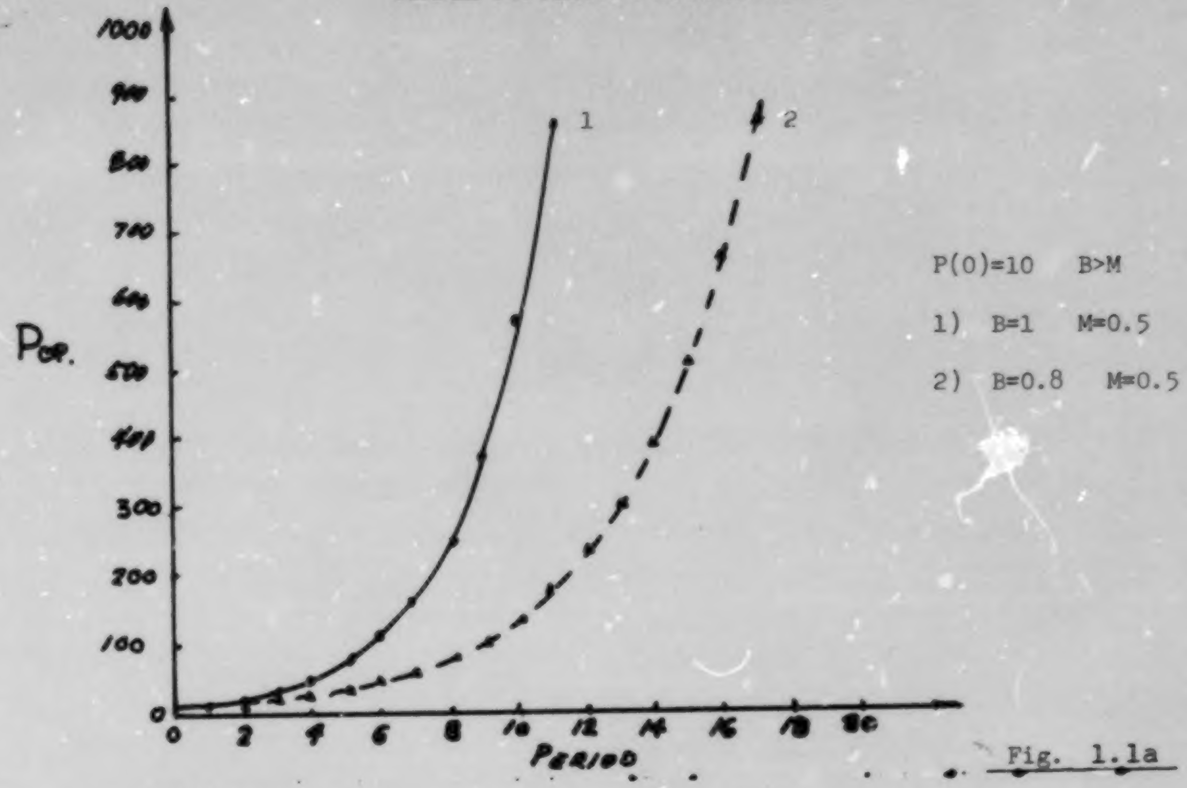


Fig. 1.1a

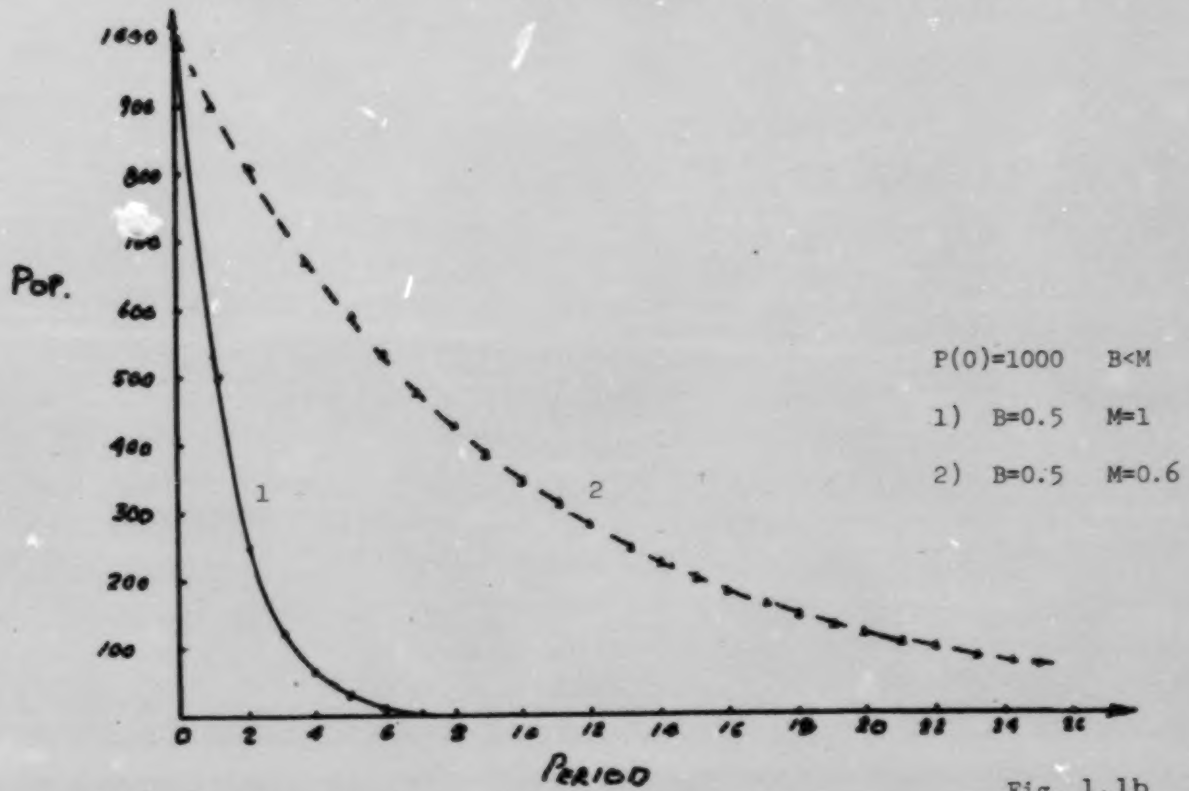


Fig. 1.1b

quite as useless as we may have initially thought. Similarly, varying the magnitude of (B-M) has the effect of changing how fast the population grows or dies out. Moreover, we see that it was the difference in the population at adjacent time periods that gave us helpful insight and so we should alter our program to print or plot this information.

It is important that the student recognize that both of the coefficients, B and M, are proportions per unit time. This can be more easily understood by considering a numerical example. Until recent years the birth rate in the United States was about 2.2%. This means that in a period of time of one year, approximately 2.2 children were born for each 100 persons alive during that year. It is also possible to speak of a monthly birth rate and it would be a little less than twelve times the previous figure because of the effect of "monthly compounding". Similarly, it is possible to give a birth rate per decade and this would be a little more than ten times the yearly rate. (See Chapter IV). In this way, time is accounted for in our model. For those students who have had a calculus based ecology course, the term natural or intrinsic growth rate is used. The relation between the calculus and our approach is briefly discussed in the appendix to this chapter and elsewhere throughout the text.

The constant environment problem resulted in a model whose solution was termed exponential growth or exponential decay. Because many scientific phenomena behave like exponential growth or decay, these concepts are very useful. This fact is mentioned because we wish to emphasize that exponential-like behavior of a variable is the direct result of assuming that the rate of increase or decrease of the variable is directly proportional to the magnitude of the variable. In this work, whenever such an assumption is made concerning the phenomena under consideration, we shall use the term exponential growth or exponential decay to describe the resultant behavior. Thus, any variable satisfying an equation of the form

$$Y(I+1) = Y(I) \pm K*Y(I)$$

where K is a constant, will be said to be exponential.

This equation, when written in the form

$$Y(I+1)-Y(I)=IK*Y(I)$$

states that the change in the variable is proportional to the variable. The idea of stating that the change in a variable is proportional to the variable or proportional to functions or combinations of other variables is extremely useful in formulating governing equations. In fact, nearly all of the examples in this work are formulated on this basis. This notion is emphasized because it frequently will enable the student to more readily construct or devise the governing equations.

This page missing from document.

Best Copy Available

facts suggest the following changes in the coefficients of natality and mortality:

Replace B by the quantity B-B1*P(I)

and

replace M by the quantity M+M1*P(I).

Thus, as P(I) increases the terms B1*P(I) and M1*P(I) will also increase and so the quantity B-B1*P(I) will decrease whereas the quantity M+M1*P(I) will increase. The student should note that in our original model the proportion of births, B, and the proportion of deaths, M, were constant throughout the problem whereas now these proportions are different each time period. This notion of varying the proportions each time period is a very fruitful notion and will be used extensively in the subsequent developments. These considerations then imply that in an arbitrary time period the change in population is given by

```
(B-B1*P(I))*P(I)-(M+M1*P(I))*P(I)
```

or

```
((B-M)-(B1+M1)*P(I))*P(I).
```

Thus, line 40 in our Malthus growth program should read:

```
40  LET P(I+1)=P(I)+((B-M)-(B1+M1)*P(I))*P(I).
```

A more concise expression, requiring fewer numerical operations, can be obtained by introducing the notation G1=B1+M1 and rewriting line 40 as

```
40  LET P(I+1)=P(I)+(G-G1*P(I))*P(I).
```

Statement 20 of the Malthus program should also be altered to read

```
20  INPUT B, M, B1, M1, P(0)
```

and a statement must be inserted to calculate G1. G1 is called the auxilliary growth coefficient.

The procedure of altering a former program in accord with alterations of the assumptions upon which the former program was constructed will be repeatedly followed in this work. The ease with which the consequences of the assumption of a finite environment were effected in the program should be noted. Nearly all of the original Malthus model program was usable. The imposition of the altered hypotheses required only minor changes and the addition of one programming statement. This is an example of the great ease with which a computer can be used to assist the obtaining of insight and understanding. A copy of the program is shown below.

```
 1  REM   FINITE RESOURCE MODEL
10  DIM P(60)
20  INPUT B,M,B1,M1,P(0)
25  LET G=B-M
26  LET G1=B1+M1
30  FOR I=0 TO 50
40  LET P(I+1)=P(I)+(G-G1*P(I))
50  NEXT I
60  FOR I=0 TO 50
70  PRINT I,P(I)
80  NEXT I
90  END
```

An examination of statement 40 shows that the essential parameters are the growth coefficient G and the auxiliary growth coefficient G1. Thus, the program could have been simplified by replacing line 20 with

```
20  INPUT G, G1, P(0)
```

and deleting lines 25 and 26. This reduced form of the program more closely approximates the forms seen in the literature. By running the program with different sets of constants B, M, B1 and M1, that is, different sets of values for G and G1, the student will note that the populations begin as exponential-like growth, reach a maximum rate of growth, then level off and approach a limiting value.

The limiting value attained by the population will be denoted by P_f. Figure 1.2, page 1.12, presents two curves obtained with two typical sets of constants. The curves are the familiar logistic or S curves that biologists frequently speak of. Problems (1) and (2) contain important results about this program.

Problems:

(1) By examining the change in population in a time period, derive an expression for the limiting population in terms of the constants B, M, B1 and M1.

(2) Make some computer runs varying just the initial population P(0). From an analysis of these runs, what can you conclude about the final population in each case? How could you determine this from your program?

The student will note that changing the constants alters the shape of the curve. For instance, the curve may begin flatter or steeper, rise more quickly or slowly and level out sooner or later depending upon what selection of constants is used. In a subsequent chapter, we will discuss how this dependency of the shape of a curve on the constants may be used to provide a method for their determination.

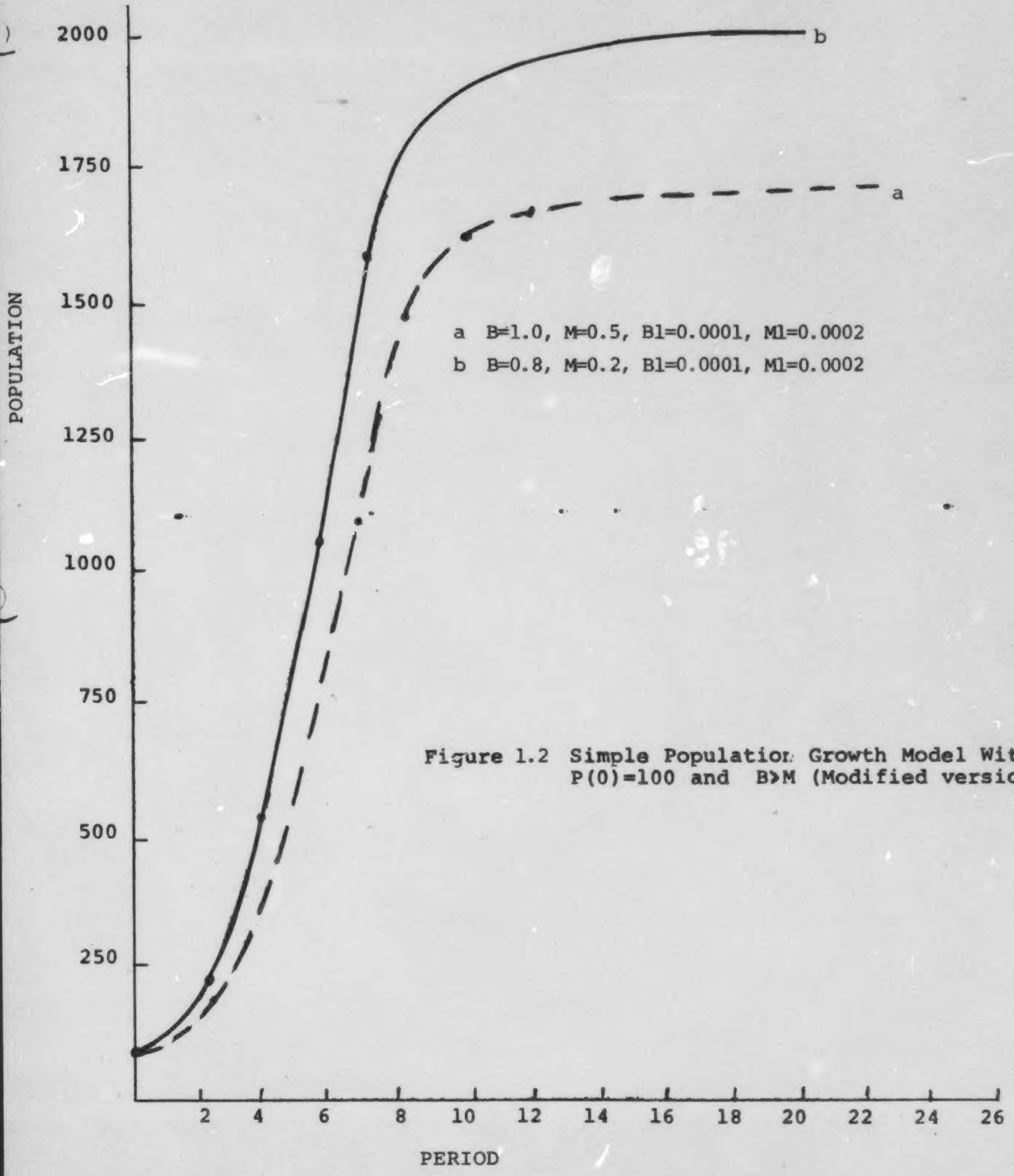


Figure 1.2 Simple Population Growth Model With P(0)=100 and B>M (Modified version)

It is interesting to note that this curve does not increase without bound as the curves obtained from the former model did. Moreover, the shape of the curves has been drastically changed as the first part of the graph curves up while the latter part does not. This is due to the changing of the coefficient of P(I). The student should rerun the original or exponential growth model and change the constants B and M after 10 or 20 time periods and then note the alteration in the resultant curve. The sudden change in the slope of the curve might suggest to the student that a frequent and/or continual changing of the coefficients B and M would enable the model builder to chang the shape of the curve so that it is "more reasonable" (whatever that is). Thus, the student could have discovered, by just "playing around with" the original model, the alterations we suggested above. Of course, if he is fortunate enough to find an alteration which more closely approximates that which actually takes place, he is faced with the problem of rationally explaining his alteration. Strange as it may seem, quite frequently this very process has resulted in better explanations of a scientific phenomena. Thus, the computer becomes an actual aid to acquiring a better understanding.

The simple relation $P_f = G/G1$ (See problem 1) permits an easy experimental determination of the constant G1. P_f may be determined by measuring the final population observed in an environment whose total food supply is constant. Since G is assumed known, G1 is given by

$$G1 = G/P_f.$$

P_f is called the carrying capacity of the environment.

Contamination of the Environment

Every student is aware of the pollution and contamination of our environment due to the expanding needs of our increasing population. The causes and effects of contamination are under extensive investigation and are becoming better understood due to the deepening concern of society with the preservation of our environment. Because of the many kinds of pollution and the diversity and multiplicity of the effects of each pollutant, and combination of pollutants, it will not be possible in a simple way to include the effects of each pollutant on the growth of the population. It is certainly true that a changing population greatly influences the growth and changes in the kinds of pollutants which in turn affect the growth of the population. The phenomenon of the growth of the population affecting the growth of the contaminants and these in turn affecting the growth of the population is an example of "feedback". This is a very familiar term in modeling. It is an engineering term and the idea of feedback has found extensive use in many fields. It will occur in most of our work.

The programming changes necessary to arrive at the finite resource program were alterations of the expressions for the number of births and the number of deaths in a time period for the Malthus model. This suggests that the effects of contamination or pollution may be accounted for by further altering these expressions.

A thorough inclusion of the effects of pollution in our population growth model is too difficult and so our development will be restricted to describing the evolution of a population in a very restricted environment. As an example of such an environment, we consider the growth of a bacterial culture in a finite or restricted volume when the culture is not renewed and hence the culture will gradually become intoxicated or poisoned due to the accumulation of catabolic products.

The following rather intuitive thoughts about the intoxication of the culture by the bacteria and the consequent effect of this intoxication on the growth of the bacteria will serve as an aid to an attempt to include the effects of intoxication in a bacterial growth model. It is certainly clear that in order to estimate the effect of an intoxicant on the change in population in a time period that the amount of the intoxicant present during that time period must be known. Furthermore, since it is assumed that the culture is not restored nor altered in any way during the time of growth of the bacteria, it seems reasonable to further assume that an amount of intoxicant present at one time period will continue to be present for all succeeding time periods. Thus, an amount of intoxicant created in a previous time period will continue to have a deleterious effect upon the population for all succeeding time periods. Hence, the total intoxication present at any time period is the accumulation of the intoxicants created during each preceding time period. The more mathematically trained student will recognize that this hypothesis merely states that the total amount of the intoxicant is a time integral of the rate of creation of the intoxicant.

In order to estimate T(I), the amount of intoxication present at the beginning of the I^{th} time period, we will suppose that the amount of intoxication created in a time period is proportional to the existing population. If C denotes the contamination constant of proportionality, then the amount of contaminant created in the I^{th} time period is C*P(I-1). Since the amount of intoxicant present in the I^{th} time period is the sum of the amounts of intoxicant produced in the previous time periods, we can write

T(I)=C*P(0)+C*P(1)+C*P(2)+.....+C*P(I-1).

Now the total intoxication should have a deleterious effect upon the population by decreasing the proportion of births and increasing the proportion of deaths. We will assume that in a time period the decrease in the proportion of births due to such intoxication is proportional to the total intoxication present in the time period and B2 will denote the constant of proportionality. Similarly, it is postulated that, in a time period, the increase in the proportion of deaths is also proportional to the total intoxication present in the time period and M2 will denote the constant of proportionality. The effect of intoxication may then be accounted for by altering the birth and death proportions to read:

B-B1*P(I)-B2*T(I) and M+M1*P(I)+M2*T(I).

Hence, in an arbitrary time period the change in population is given by

(B-B1*P(I)-B2*T(I))*P(I)-(M+M1*P(I)+M2*T(I))*P(I)

In this expression, the first term represents the change in population due to births in the time period and the second term is the change in population due to deaths in the same period.

Upon introducing the notation, $G2=B2+M2$, line 40 in our program should now read

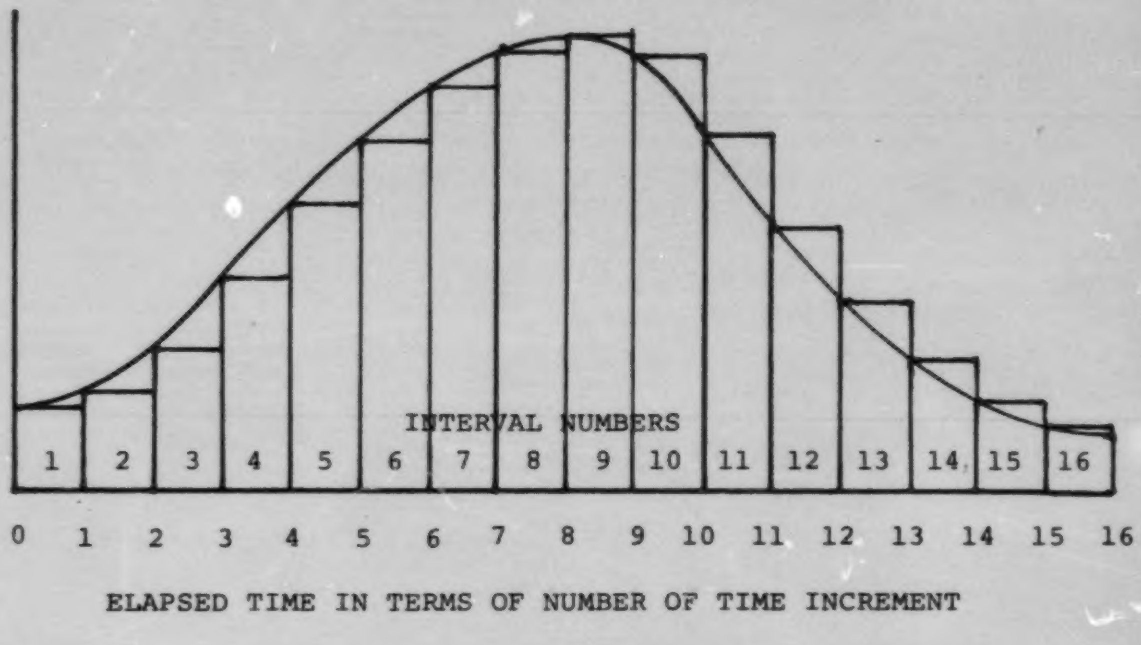
```
40  LET P(I+1)=P(I)+(G-G1*P(I)-G2*T(I))*P(I).
```

In addition, the program must be altered to provide for the new constants B2 and M2, the calculation of G2 and the total intoxication T(I).

Retarded Time Effects

The previous development assumed that the effect of contamination was independent of the time of creation or deposition of the contaminant. This is a restrictive hypothesis since in a completely enclosed environment, such as a microbial population growing in a culture, it is known that the effect of the contaminant varies according to the age of the contaminant. Thus, intoxicants created early in the history of the culture have a different effect than intoxicants recently created. The inclusion of such a retarded time effect is accomplished by modifying the calculation of the potency of the contaminant.

The discussion of the development of a computer program which includes the effect of the time of existence of the contaminant is facilitated by an examination of a graphical portrayal of a typical population curve such as is shown in figure 1.2a.



A TYPICAL POPULATION CURVE

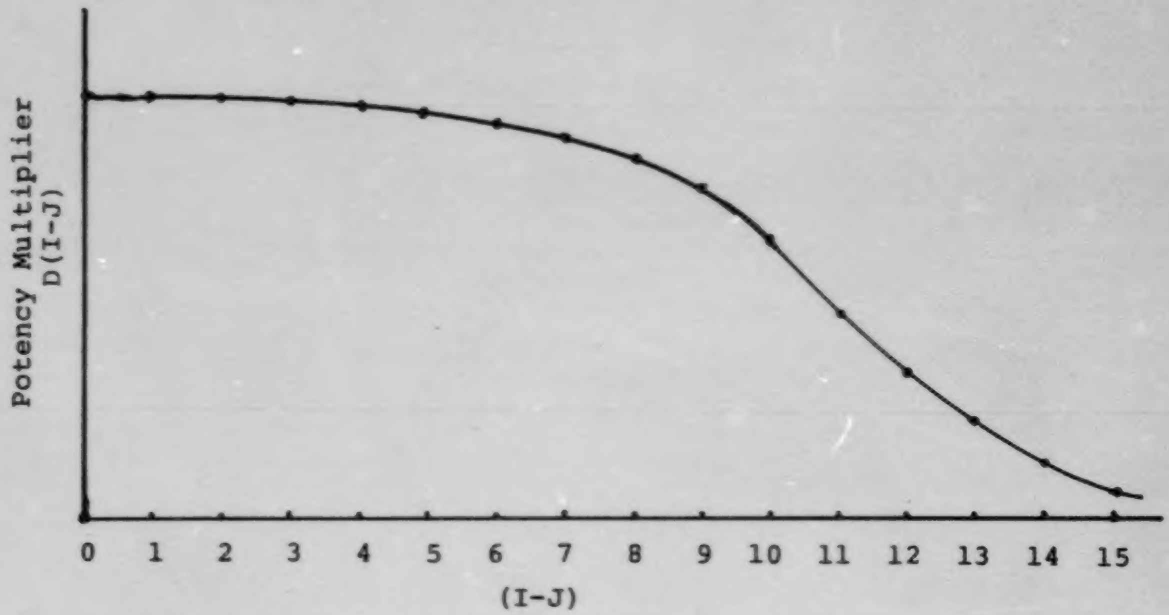
Fig. 1.2a

The population is shown as a sequence of steps in recognition of the fact that the population is actually calculated at a discrete set of points. It is assumed that the time increments are constant and that the time index I, indicates the elapsed time measured in numbers of time increments. As an example, if the time increment is one hour, the value I=14 indicates that the elapsed time is 14 hours. We will also adopt the convention of denoting the initial time by 0, and hence P(0) is the initial population. For this reason, P(I) will denote the population at the end of time t=I time increments. These conventions mean that if we are calculating the population P(I+1) from a knowledge of the system up to and including time t=I time increments, then we are indeed calculating the change in the population during the (I+1)st time period. It will further be assumed that the amount of contamination created during a period is proportional to the population at the beginning of that time period. Thus, the contamination created in the J^{th} time period is due to P(J-1). As a specific example, the population at the end of the 6^{th} generation, P(6), creates an amount of contamination equal to C*P(6) during the 7^{th} time period.

Using this notation we can begin our discussion of the development to include retarded time effects in our growth model. We start by estimating the potency or degree of intoxication due to the creation of an amount of contaminant during an earlier time period. If the early time period is the J^{th} time period, then according to our original hypothesis, the amount of contaminant created in this period is C*P(J-1). To calculate the length of time that the contaminant has been in existence it is helpful to recall that we are at that point in the program where we have calculated the population P(I) and are attempting to calculate the population one time increment later, that is P(I+1). We will adopt the further convention that the intoxication created in a time interval is not deposited until just before the end of the time interval. This implies that the contaminant created in a time period has no effect on the population growth during that time period. Such a convention also implies that the length of time since deposition of the contaminant is given by the difference between the present generation number, I, and the time period, J, in which the intoxicant was created. Now, because it is assumed that the potency of the contaminant depends upon the elapsed time since deposition of the contaminant, the expression

$$D(I-J)*C*P(J)$$

represents the toxicity or deleterious effect during the (I+1)st time period of the amount of intoxicant deposited in the (J+1)st time period. D(I-J) is a coefficient or multiplier which accounts for the potency of the contaminant due to the time of existence of the contaminant relative to the potency of an equal amount of contaminant which has just been created. For example, if the time period is one hour and we are calculating the change in the population during the 11th hour, D(10-5) is the effect on the potency of the contaminant due to the fact that the contaminant has been in existence for five hours. Furthermore, the assumption that the contamination created in a period has no effect on the population growth during that period implies that, if for example I=4, then setting D(0)=0 assures that the amount of contaminant created during the 5th time increment, C*P(4), has no effect on the calculation of P(5). The coefficients D(I-J), (I-J)=0, 1 2, ,,,,, I, must be entered as data or else provision must be made for their generation. A graphical representation of one possible form of the time variation of D(I-J) is given in figure 1.2b.



Time Since Deposition Measured in Number of Time Increments
Fig. 1.2b

The student should note that increasing times since deposition correspond to earlier times of creation of the contaminant. Because of this, the student should be careful in the interpretation of the horizontal time scale. In essence, increasing values of (I-J) link the present generation with generations further back in time.

The toxicity T(I), of the total amount of the contaminant present during the (I+1)st time interval is given by the expression

D(I)*C*P(0)+D(I-1)*C*P(1)+D(I-2)*C*P(2)+..... + ... +D(I-J)*C*P(J)
+...+D(1)*C*P(I-1)+D(0)*C*P(I).

The programming equivalent of this expression is

```
120 LET T(I)=0
130 FOR J=0 TO I
140 LET T(I)=T(I)+D(I-J)*C*P(J)
150 NEXT J
```

The statement numbers serve merely to indicate the order of operations and bear no relation to their place in a particular program. In terms of summation notation of mathematics this expression may be written as

$$T(I) = \sum_{J=0}^{I} D(I-J)*C*P(J).$$

The student is urged to completely understand the programming form of the sum since this form will be used in other parts of the text. As we have repeatedly stated, we are attempting to get the student to think in terms of the BASIC language as early as possible in order that he or she may more effectively use the computer. The mathematical equivalent has been presented as a possible aid to better understanding by the more mathematically inclined students. In addition, the student should note that setting D(I-J)=1 for all J implies that there is no retarded time dependency and in this instance the above expression for T(I) reduces to that given previously.

The inclusion in the population model of this new expression for the effect of the total intoxication is accomplished by substituting the new expression for T(I) in our previous development.

A program accounting for retarded time contamination effects on a population growing for 50 generations in a finite environment is given in figure 1.2c.

In this program it is assumed that the contaminant is biodegradable and has no toxic effect after 20 time increments of existence. The degradation will be assumed to be in direct proportion to the elapsed time of deposition. Thus we will set D(1)=1, D(2)=.95, D(3)=.90, etc. down to D(20)=0. We will also set D(K)=0 for K=21, 22, ,,,, , and set D(0)=0. The assumption that D(K)=0, for K>20 implies that the con-

```
5 REM  POP. GR. MODEL WITH FIN. RES. AND ENV. CONTAMINATION
10 DIM P(60),D(60),T(60),T1(60)
30 PRINT "TYPE  G  AND  G1  THE GROWTH COEFFICIENTS"
35 INPUT G,G1
40 PRINT "TYPE  G2  AND  C  THE CONSTANTS OF PROPORTIONALITY"
45 INPUT G2,C
50 PRINT "TYPE P(0)  THE INITIAL POPULATION"
53 INPUT P(0)
54 PRINT
55 PRINT
58 REM LINES 60 TO 68 READ IN THE TIME DELAY MULTIPLIERS
60 DATA 0,1,.95,.9,.85,.8,.75,.7,.65,.6,.55
62 DATA .5,.45,.4,.35,.3,.25,.2,.15,.1,.05
64 FOR K=0 TO 20
66 READ D(K)
68 NEXT K
70 FOR I=0 TO 50
74 REM  LINE 75 ASSUMES NO INITIAL CONTAMINATION
75 LET T(I)=0
80 FOR J=0 TO I
90 LET T(I)=T(I)+D(I-J)*C*P(J)
100 NEXT J
104 REM  LINE 105 STORES THE T(I)'S FOR OUTPUT IN LINE 150
105 LET T1(I)=T(I)
110 LET P(I+1)=P(I)+(G-G1*P(I)-G2*T(I))*P(I)
113 IF P(I+1)<0GO TO 170
120 NEXT I
130 PRINT " I                P(I)           T1(I)"
131 PRINT
140 FOR I=0 TO 50
150 PRINT I,P(I),T1(I)
160 NEXT I
165 GO TO 200
170 PRINT "THE POPULATION BECAME NEGATIVE"
200 END
```

Single Population Growth Assuming a Finite Resource Environment and Retarded Time Contamination Effect

Fig. 1.2c

taminant completely degraded and therefore has no effect after 20 time increments and the setting of D(0)=0 assures that the contaminant created by P(I) does not affect the growth of P(I+1). This specification of the potency multipliers is expressed by lines 60 to 68 of the program. By setting D(I-J)=1 for all values of (I-J), the original contamination model is obtained and by setting D(I-J)=0 for all values of (I-J), the original finite resource model is obtained. These choices of values for D(I-J) are useful for debugging the program because they enable a comparison with the results of previously developed programs. Instruction 110, which calculates the population for the beginning of the (I+1)st time increment, requires a knowledge of the potency of the contamination during the I^{th} time increment. This is accomplished by statement 90, which calculates the total potency to be used in the I^{th} time interval. Instruction 113, is necessary to assure biologically realizable results. Statements 30 and 35 call for only the four values G, G1 and G2, as well as C, rather than the seven values, C, B, M, B1 M1, B2 and M2. This results in no loss of generality since the latter six values are combined to give the former three values prior to the basic population calculation.

The obtaining of "reasonable" values for the parameters requires some deliberation, as well as experimentation with the program. The student will recall that G1<<<G since both B1 and M1 were very much smaller in magnitude than B and M. This was due to the fact that the effect of a finite environment was to modify or alter the birth and mortality rates. In a similar manner, it is assumed that the effect of contamination will be to only further alter the birth

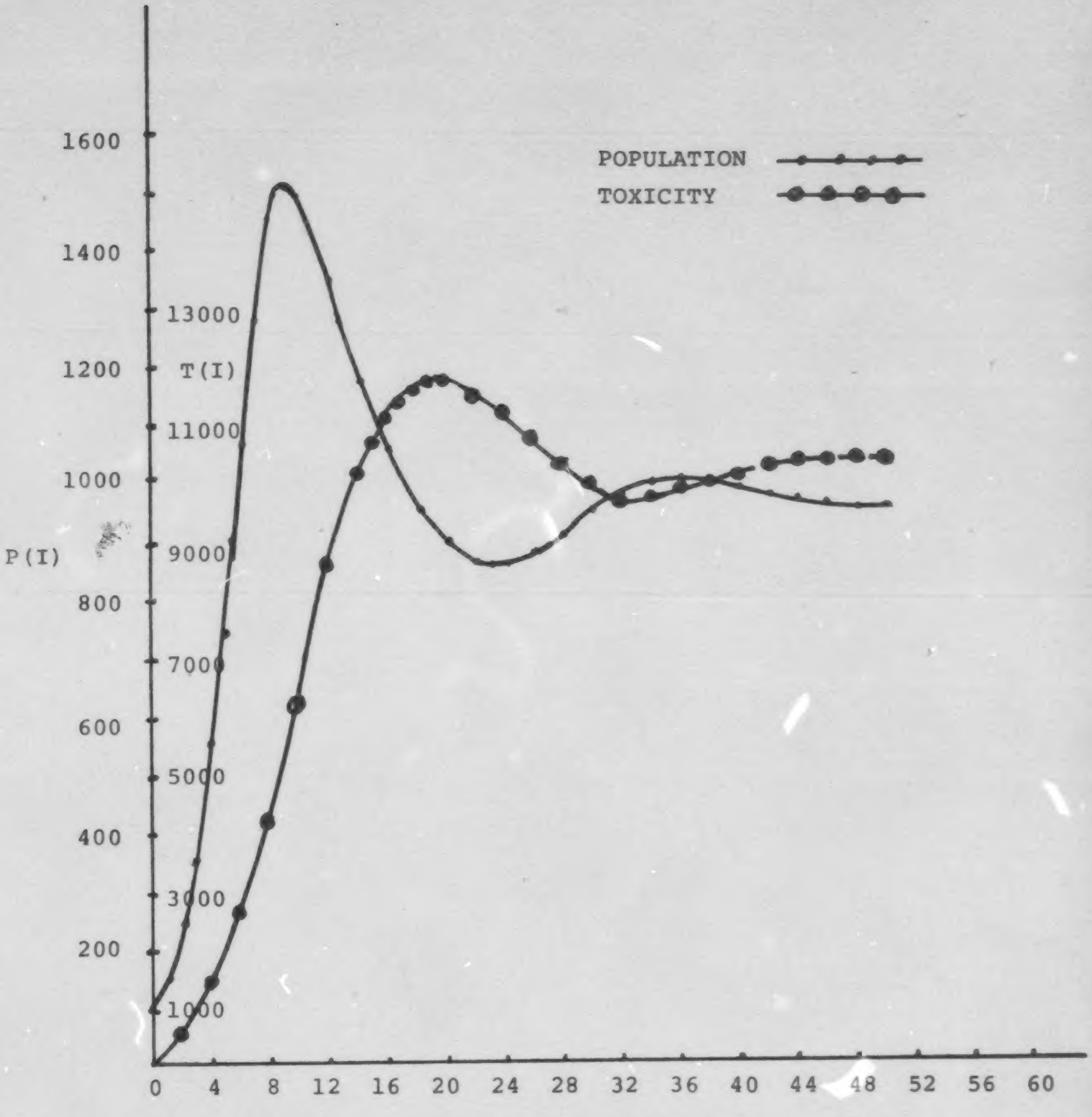
and mortality rates. Thus, the terms (B-B1*P(I)) and (M+M1*P(I)) still remain the principle contributions to the birth and death rates. Hence, it is expected that G2, the constant of proportionality modifying the growth rate because of contamination effects, will be much smaller in magnitude than either G or G1. The presence of the constant C is actually not required since an examination of the program shows that the effect of C can be incorporated into the magnitude of G2. For this reason, we usually set C=1. C appears in the program, however, because it is occasionally the case that it is desirable to change all of the multipliers by a constant proportion and such a change can be readily accomplished by only altering C.

The student should note that the values chosen for the multipliers in the program are arbitrary. The student is encouraged to try other sets of values and to compare the results. It is also instructive to compare results obtained from the program when different magnitudes of G2 relative to the magnitude of G and G1 respectively are used. Your author tried the combination G=0.6, G1=0.0003, G2=0.00003 and C=1.0. The program results are listed in figure 1.2d and displayed in graphical format in figure 1.2e. The latter figure shows that the early time history of the population is exponential in character. As the toxic effect acquires sufficient magnitude, the population is then decreased. This decrease in population eventually results in a decrease in the magnitude of the toxicity which in turn permits

```
TYPE  G  AND  G1  THE GROWTH COEFFICIENTS
? 6, .0003
TYPE  G2  AND  C  THE CONSTANTS OF PROPORTIONALITY
?.00003, 1
TYPE P(0)  THE INITIAL POPULATION
?100


 I             P(I)          T1(I)

 0            100           0
 1            157           100
 2            243. 334      252
 3            369. 732      482. 484
 4            545. 209      827. 199
 5            769. 628      1328. 9
 6            1023. 02      2027. 77
 7            1260. 63      2941. 55
 8            1429. 01      4041. 78
 9            1500. 52      5247. 36
 10           1489. 15      6453
 11           1429. 09      7572. 25
 12           1349. 21      8556. 97
 13           1266. 27      9390. 37
 14           1188. 28      10073. 4
 15           1118. 55      10615. 1
 16           1058. 13      11027. 6
 17           1007. 06      11323. 8
 18           964. 93       11516
 19           931. 197      11615. 7
 20           905. 28       11633. 5
 21           886. 641      11578. 8
 22           874. 799      11465. 2
 23           869. 205      11303. 2
 24           869. 328      11104. 1
 25           874. 612      10880. 2
 26           884. 418      10645. 3
 27           897. 964      10414. 9
 28           914. 274      10205. 1
 29           932. 162      10029. 7
 30           950. 302      9897. 9
 31           967. 381      9812. 68
 32           982. 284      9771. 47
 33           994. 239      9768. 26
 34           1002. 87      9795. 35
 35           1008. 16      9844. 66
 36           1010. 39      9908. 55
 37           1010. 01      9980. 18
 38           1007. 58      10053. 8
 39           1003. 66      10124. 9
 40           998. 8        10189. 9
 41           993. 47       10246. 4
 42           988. 072      10292. 9
 43           982. 925      10328. 7
 44           978. 268      10353. 7
 45           974. 266      10368. 3
 46           971. 024      10373. 5
 47           968. 586      10370. 4
 48           966. 951      10360. 6
 49           966. 078      10345. 6
 50           965. 893      10327. 1
```

A Typical Result From Program Listed in Fig. 1.2c

Fig. 1.2d



Population and Toxicity vs. Time
Assuming a Biodegradable Contaminant

Fig. 1.2e

the population to again increase. However, the latent toxicity present at the start of the second rise in population is such that the population does not increase to as great a value as the initial maximum population. In this way, the successive minimum and maximum of the population decrease and it is evident that if the program were run a sufficient length of time there would be no change in population from period to period. Thus, a finite and decreasing set of potency multipliers produces a population growth curve which is characterized by a rapid and large initial increase in population followed by a sequence of damped oscillations. The population finally assumed a constant value and this constant value will be called the Permitted Contamination Population.

Figure 1.3c illustrates the results of a run in which all of the potency multipliers are assumed to have the value of 1.0. Many variations in the time evolution of the population can be obtained by choosing different sets of values for the parameters and the potency multipliers. The analysis of results obtained from these variations will provide the student with insight about the behavior of the population. The examination of these variations is facilitated if a plotting routine is used to display the results in graphical form.

The student who is familiar with the integral calculus will recognize that the expression for T(I) is the discrete or finite difference equivalent of the retarded time integral

$$C\int_{\tau=0}^{\tau=t} D(t-\tau)P(\tau)d\tau$$

and thus the program we have developed numerically solves the integro-differential equation

$$\frac{dP}{dt} = [G-G1\cdot P(t)-C\int_{\tau=0}^{\tau=t} D(t-\tau)P(\tau)d\tau]P(t).$$

The mathematical study of such equations is quite difficult and is usually attempted only in graduate level mathematics courses. The term $D(t-\tau)$ is called the kernel, and, if this term is not exceedingly simple, the solution of such equations must be effected by numerical methods. Thus, even a mathematically formulated discussion of this problem would necessitate the writing of a program very similar to that which we have developed. This is an example of the power of formulating and thinking about quantitative phenomena in terms of a programming language such as BASIC. As an assignment, the student should assemble the program and then run it for several variations of the set of constants.

The model that has been constructed is relatively simple yet does produce a population curve which compares somewhat favorably with that obtained experimentally. Many known biological effects have been omitted. As an example, Allee (1938) noted that thirty specimens in a culture would neutralize over two hundred times the amount of poison normally neutralized by one specimen. This fact should certainly require a significant alteration of the early time period part of our model. The effect of overcrowding is another effect that we have omitted. The reader can very easily cite other examples of effects which have been omitted. He is urged to do so and to alter the program to account for the additional effects. The altered program should be run and the results analyzed to test his alterations and hypotheses.

In the preceding discussion, the student should note that such phrases as, "is assumed to be", "is supposed to be", "is postulated to be", etc. are to be understood as equivalent expressions. They serve to state hypotheses and the different phrases are used to avoid repetition. This technique is frequently employed in literature describing quantitative phenomena.

It is interesting to compare the results obtained from each of the three models. Figure 1.3 shows a population growth curve of a bacterial culture and is taken from the work of Buchanan and Fuller in 1928 as reported in D'Ancona (1954). Figure 1.3a is a comparison of the results obtained from a Malthus model using a positive growth coefficient. It is seen that the early time behavior of the curves is quite similar. Consequently, it may be inferred that the initial growth of a bacterial culture is exponential in character. Figure 1.3b is a comparison of the same experimental data with numerical results obtained from a finite resource model. The two curves compare quite well in shape and form up to maximum growth. For early time periods, the finite resource model agrees just as closely as does the constant environment model results. However, since the finite resource model also agrees for a longer period of time, we conclude that the latter model provides a better description of bacterial culture growth than does the former model.

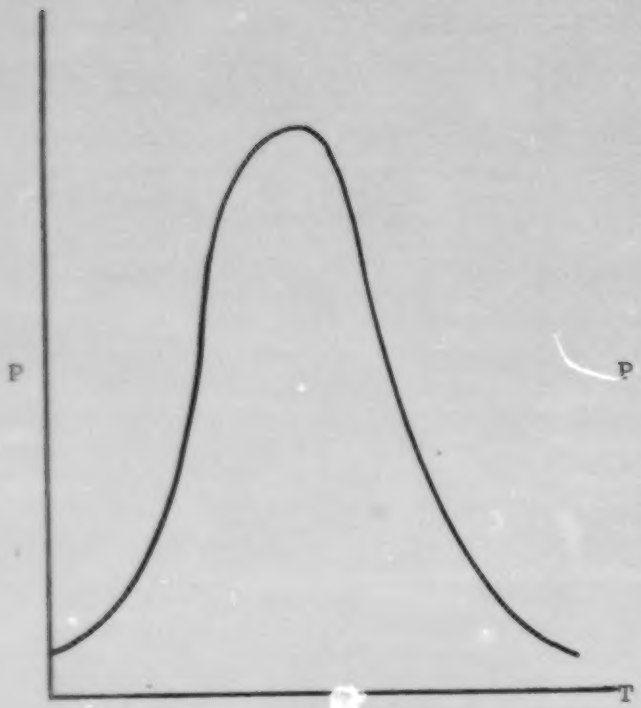


Figure 1.3

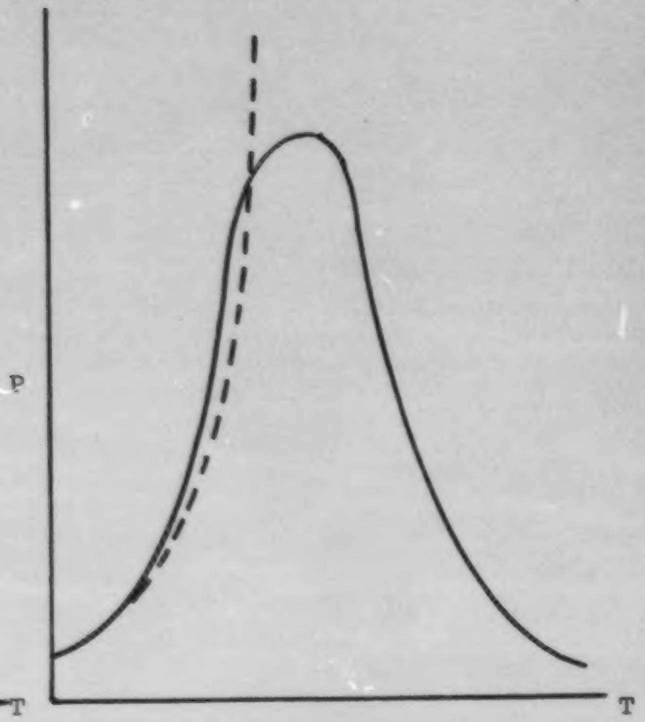


Figure 1.3a

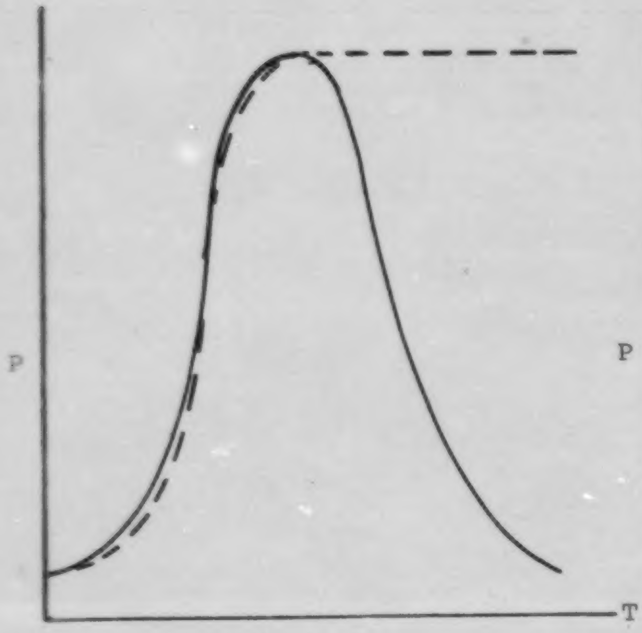


Figure 1.3b

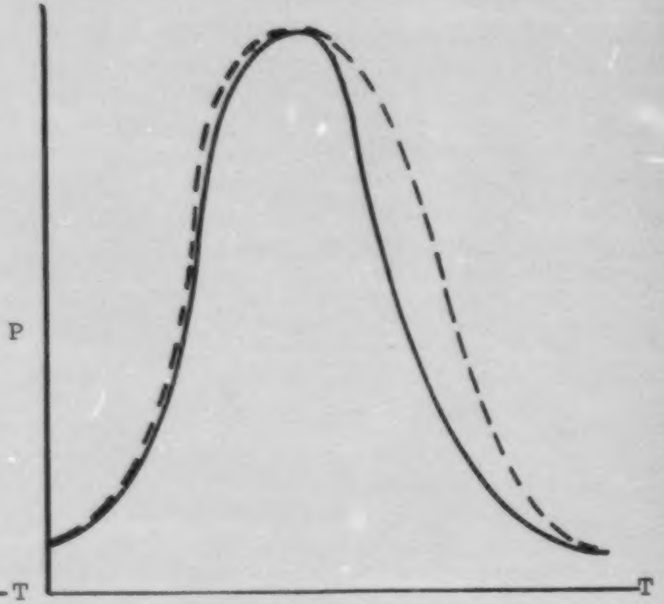


Figure 1.3c

Finally, figure 1.3c presents a comparison with the results obtained from the contamination model. The D(I-J) were all assumed to be constant and equal to 1. It is seen that, compared to the results of the other two models, the contamination model yielded results which were in longer time agreement. Hence, one may conclude that this model is an even better description of the bacterial growth.

These comparisons were made on the basis of shape and form rather than on magnitude and were made solely to illustrate how the repeated refinement of a model can yield progressively closer agreement with experimental results. In this way insight and understanding of the phenomena are increased. Chapter III describes more exact methods of comparing results. Your author is purposely avoiding positive statements to the effect that one model is definitely better than another because the former produced results which more closely agreed with experiment. It is evident that the relative worth of a model must also include an evaluation of its simplicity. It is certainly possible to construct a model which agrees more closely with experimental data by combining arbitrary functions and constants in a 'willy-nilly' manner contrived to produce closer agreement. About all that can be said is that there is no single best model. The decision as to which model is the better, is not trivial and is far better left for discussion elsewhere. The purpose in constructing models is to gain insight and understanding of the biological phenomena and comparison with experiment is necessary to confirm or to deny insight.

Effect of Mating Possibility

For populations whose members reproduce by sexual means, the frequency with which encounters take place between members of the opposite sex is a significant factor in the determination of the birth rate. If the population density is large, such encounters are frequent, whereas in a sparsely settled area such encounters can be rare. In order to simplify the analysis, we will suppose that there are a sufficient number of species present to validate the following assumptions which are due to Volterra. We assume that the proportion of sexes remains constant over the entire growth cycle and that F1 denotes the proportion of males and F2 denotes the proportion of females. Thus, since the species of the population are either male or female, F1+F2=1, and F1*P(I) and F2*P(I) are the number of males and females respectively in the I^{th} time period. We will also assume that the fraction of total possible encounters between members of the opposite sex that give rise to births is F3. Consequently, in the I^{th} time period if there are E(I) encounters, the number of resultant births is F3*E(I). Now the number of possible encounters is given by the product of the number of males and females in the period, i.e.

E(I) = (F1*P(I))*(F2*P(I))

or

E(I) = F1*F2*P(I)↑2.

Problem:

Verify the assertion that the number of possible encounters between males and females is the product of the number of females and the number of males by constructing a diagram and counting the number of encounters for the case of (a) 2 females and 3 males, and (b) 4 males and 3 females.

Consequently, the number of births in the time period is

F3*E(I) = F1*F2*F3*P(I)↑2
= F4*P(I)↑2

where we have introduced the notation F4 = F1*F2*F3. By assuming that the number of births is determined by the number of encounters of members of the opposite sex, we are in effect altering our original assumption of a constant birth rate B. Recall that a constant birth rate implied that B*P(I) was the number of newborn in the I^{th} time period. This expression must now be replaced by the expression F4*P(I)↑2. A comparison of this expression with B*P(I), suggests that the constant B must be replaced by F4*P(I). Since we are modifying the program developed in the previous section to account for the effect of mating possibility, and since G = B-M, we must write G = -M+F4*P(I) and the coefficient of P(I) in line 40 should read

```
-M+(F4-G1)*P(I)-G2*T(I).
```

Thus, line 40 page 1.16 should read

```
40  LET P(I+1) = P(I)+(-M+(F4-G1)*P(I)-G2*T(I))*P(I).
```

The assumption that the proportion of encounters resulting in births remains constant is a crude assumption at best. In the early stages of the population growth, this assumption might be reasonable but as the population gets larger the assumption becomes invalid. It seems more reasonable to assume that as the population increases the proportion of births due to encounters should decrease. Thus, we replace the constant F3 by F3-F5*P(I) where F5 is positive and so small that the quantity F3-F5*P(I) remains positive during the entire growth cycle. The number of births due to the encounters E(I) is then

```
(F3-F5*P(I))*E(I).
```

Letting F1*F2 = F6, permits us to write this expression as

```
(F3-F5*P(I))*F6*P(I)↑2.
```

Line 40 then may be written as

```
40  LET P(I+1) = P(I)+(-M+(F3-F5*P(I)-G1)*P(I)-G2*T(I))*P(I).
```

Again, the student should provide the necessary alterations to his existing program and carry out a few runs with various sets of the parameters M, B1, B2, M1, M2, F1, F2, F3, F5 and P(0). In summary, equation 40 above includes the three effects (1) a finite food supply, (2) contamination, and (3) mating.

Summary

This chapter has considered the general problem of the growth of a single population. The development began with a consideration of the growth of a population in a constant environment and successive alterations of the fundamental hypotheses resulted in a set of models each of which more closely mimiced reality. Because the alterations consisted in modifications of C(I), and the addition of input statements, their accommodation was readily accomplished. Thus, this procedure enabled an easy and ready transition from one model to another.

This is in contrast to the mathematical procedure which usually always requires the development of a new method of solution whenever the original model is altered. Furthermore, the development of such methods is not trivial and can be a very difficult and time-consuming task. In fact, it is usually the case that far more time is spent developing a method of solution to a particular problem than is spent discussing the solution of the problem or in discussing the correctness of the original formulation of the problem. The programming language approach permits the scientist to concentrate on learning more about the phenomena under investigation rather than on the learning of some mathematical technique which is usually applicable to only a very specific and restricted class of problems.

The direct BASIC programming language approach also permits the investigator to place far greater emphasis on obtaining a more complete and correct formulation of the problem together with a more extensive analysis of the results. This approach is in direct contrast to the applied mathematics approach which has sometimes been characterized as the art of linearization or the art of simplifying the problem to the extent that the resulting mathematical expression of the problem is solvable, yet the essence of the actual phenomena is not lost.

A further value of the direct programming approach is the ease with which it permits the investigator to examine the effects of various biological hypotheses. This, in turn, enables the investigator to obtain a better understanding of the overall structure, as well as the overall interaction of the various components making up the system. In this way, major weaknesses in the system are more easily uncovered.

APPENDIX

Malthus or exponential growth is frequently described in the language of the calculus. It is the purpose of this appendix to indicate, in a most informal manner, the connection between the two approaches. The notion of the instantaneous time rate of change of a variable or function is a fundamental concept in the calculus. Heuristically speaking, the instantaneous time rate of change of a function is the change in the function in a very small increment of time. If dt denotes a small increment of time and dF denotes the change in the function F during the increment of time, then the quotient, $\frac{dF}{dt}$ is a close approximation to the instantaneous time rate of change of the function. In the calculus, the value of $\frac{dF}{dt}$ as the increment of time becomes vanishingly small is made precise, and this value is called the derivative of the function with respect to the time. Now, the Malthus hypothesis states that the time rate of change of the population is proportional to the population. Hence, we can write

$$\frac{dP}{dt} = rP \qquad (1)$$

where r is the constant of proportionality and is called "the intrinsic rate of growth". The analogy to the BASIC language formulation of the problem can be noted by recalling that in the derivation of statement 40 in the Malthus program it was assumed that the interval of time was a unit interval, i.e. a single generation, a single year, a single day, etc. If the time interval had been chosen to be H units long, the index I would have been related to the actual time by

$$T = I * H \qquad (2)$$

Thus, if B and M are interpreted as the respective birth rate and the mortality rate per unit time, the proportion of births and deaths in a single period is B*H and M*H respectively. The fundamental BASIC language equation is then

```
P(I+1) = P(I)+B*H*P(I)-M*H*P(I)
```

or

$$P(I+1) = P(I)+G*P(I)*H \tag{3}$$

where $G = B-M$.

By setting $H=1$ statement 40 in the program listed on p. 1.6 is obtained. The analogy with the term $\frac{dP}{dt}$ can be noted by writing equation (3) in the form

$$\frac{P(I+1)-P(I)}{H} = G*P(I) \tag{4}$$

since it is evident that the lefthand side of this equation is the change in the population during the time H; that is, the lefthand side is the time rate of change of the population. Thus, the term

$$\frac{P(I+1)-P(I)}{H}$$

is an approximation to $\frac{dP}{dt}$. From equation (2) it is seen that the term $P(I+1)-P(I)$ is the difference in populations occurring in a time period whose duration is H units of time. Thus, if H is made successively smaller the value of

$$\frac{P(I+1)-P(I)}{H}$$

approaches the value of the instantaneous time rate of change of the population, $\frac{dP}{dt}$.

The relation between the growth coefficient and the intrinsic rate of growth is readily determined. From the calculus it is known that the solution to equation (1) is

$$P(t) = P(0)e^{rt} \tag{5}$$

where $P(0)$ is the initial population. By writing equation (4) in the form

$$\frac{P(I+1)-P(I)}{P(I)} = G*H \tag{6}$$

and substituting expression (5) we obtain

$$\frac{P(0)e^{r(I+1)H} - P(0)e^{rIH}}{P(0)e^{IrH}} = G*H$$

or

$$e^{rH} - 1 = G * H \tag{7}$$

If H=1, this simplifies to

$$e^{r} - 1 = G \tag{8}$$

which is a form that is frequently presented.

It is the case that the graphical display of data which has a large variation is more easily accomplished by using a compressed scale. A typical example of data having a wide variation is exponential growth data. To facilitate the graphical display or plotting of such data, compression by a logarithmic scale is usually employed. Logarithmic plotting is accomplished by plotting the logarithm of the function values rather than the actual value of the functions. However, to avoid the necessity for the use of antilogarithms when reading the data from the graph, the vertical axis is usually labeled with the actual values of the function. Since the logarithm subroutine requires several arithmetic operations, the frequent use of this routine could require an excessive amount of computer time. Consequently, it is desirable to have an alternative but "cheaper" method for compressing the data. It is the purpose of the following discussion to present such an alternative by developing an operation which is analogous to the logarithm operation and which for sequential data requires far fewer arithmetic operations than does the logarithmic subroutine.

We begin by considering an arbitrary smooth function G. The expression

$$\frac{dG}{dt} \times \overline{\Delta t}$$

is the rate of change of the function G multiplied by a small increment of time $\overline{\Delta t}$, and the product is the change in G during the elapsed increment of time. Thus, if $\overline{\Delta t}$ denotes the difference in time between the I^{th} time increment and the (I+1)st time increment we have

$$\frac{dG}{dt} \times \overline{\Delta t} = G(I+1)-G(I)$$

Since $\frac{dG}{dt}$ varies in time, the successive products of $\frac{dG}{dt}$ with small increments of time gives the change in G in each of the successive time increments. If $\overline{\Delta t}_I$ denotes the I^{th} time increment and $\left.\frac{dG}{dt}\right|_I$ denotes the rate of change of G in the I^{th} time increment, the sum of the changes in G in N successive time increments is

$$\left.\frac{dG}{dt}\right|_1 \overline{\Delta t}_1 + \left.\frac{dG}{dt}\right|_2 \overline{\Delta t}_2 + \ldots + \left.\frac{dG}{dt}\right|_I \overline{\Delta t}_I + \ldots + \left.\frac{dG}{dt}\right|_N \overline{\Delta t}_N$$

$$= (G_2-G_1)+(G_3-G_2)+ \ldots +(G_I-G_{I-1})+ \ldots +(G_N-G_{N-1})$$

$$= G_N-G_1 . \qquad (9)$$

Here G_I denotes the value of G at the end of the I^{th} time increment.

In the calculus it is shown that

$$\frac{1}{F}\frac{dF}{dt} = \frac{d}{dt}\ln F \qquad (10)$$

where lnF denotes the natural logarithm of F. Furthermore, it is evident that the term

$$\frac{F(I+1)-F(I)}{H\ F(I)} \qquad (11)$$

is an approximation to the term $\frac{1}{F}\frac{dF}{dt}$, and, because of equation (10), is also an approximation to the time rate of change of the natural logarithm of F. In equation (9) if we set G=ln F, we have

$$\frac{d}{dt}\ln F\bigg|_1 \overline{\Delta t}_1 + \frac{d}{dt}\ln F\bigg|_2 \overline{\Delta t}_2 +\ldots+ \frac{d}{dt}\ln F\bigg|_I \overline{\Delta t}_I +\ldots+ \frac{d}{dt}\ln F\bigg|_N \overline{\Delta t}_N$$

$$= \ln F(N) - \ln F(1). \qquad (12)$$

This result suggests that the sum

$$S(N)= \frac{F(2)-F(1)}{H\ F(1)}\ xH + \frac{F(3)-F(2)}{H\ F(2)}\ xH +\ldots+ \frac{F(I)-F(I-1)}{H\ F(I-1)}\ xH +\ldots+ \frac{F(N)-F(N-1)}{H\ F(N-1)}xH \qquad (13)$$

is analogous to the quantity

$$\ln F(N) - \ln F(1).$$

This sum is the desired operation on F that will be used to compress the scale of F. Whenever F is generated in sequential fashion, the calculation of S(N) for any step involves only a subtraction, a division and an addition. For this reason, the calculation of S(N) can be inexpensively carried simultaneously in the program. Two examples are given below.

The first example is the Malthus population growth problem. The program listed on page 1.6 is modified to calculate S(I) in accordance with the above work. The modified program is listed on page 1.33. A sample run was made assuming that the initial population was 10 individuals and the growth coefficient was 0.2.

A graphical portrayal of the results is given in figure 1.4. The straight line graph is analogous to the straight line plot obtained when graphing Malthus type growth using a logarithmic scale. By comparing this plot with the conventional portrayal in figure 1.2, the student will appreciate the advantages of each mode of plotting.

A second example is the growth of a population restricted by a finite resource. The program and the results of a typical run are listed in figures 1.5a and b. In figure 1.5b, the first column indicates the number of the time period and the second column lists the population at the beginning of the time period. The third column indicates the "compressed" value of P(I) and the last column is the natural logarithm of the difference between the present population and the initial population. This provides a measure of the data compression of each operation. By comparing the values of S(I) and P(I) it is seen that considerable data compression is effected. A comparison of columns (2) and (4) illustrates logarithmic data compression. Statements 45 and 47 calculate the sum indicated by equation (13) and statement 48 calculates the natural logarithm of the population.

In figures 1.4 and 1.5a, the actual population values are listed on the vertical axis to facilitate the reading of the data. Because such a labeling of the vertical scale is non-linear, the accurate reading of the plotted data is not possible. It is usually the case, however, that data plotted in this manner is not plotted for the purpose of permitting accurate determination of the function. Rather, the purpose of such plotting, is to display the data so that the overall behavior of the function can be determined and analyzed.

The student who is familiar with the calculus will note that the development leading to equation (10) was actually a cavalier derivation of the fundamental theorem of the calculus since equation (9) is a very close approximation to the integral of the rate of change of a function. Your author wants to emphasize that these portions of the discussion relating to the derivative and the integral are very heuristic. Nevertheless, they can be made rigorous and are made so in the calculus. It also should be emphasized that S(N), as calculated, is only an approximation to the difference of the natural logarithms of the final

```
1 REM     CONSTANT ENVIORNMENT; COMPRESSED SCALE CALCULATION
2 REM
3 REM
10 DIM P(60),S(60)
15 PRINT "TYPE  G,  P(0)"
20 INPUT G,P(0)
25 LET S=0
30 FOR I=0 TO 49
40 LET P(I+1)=P(I)+G*P(I)
45 LET R=(P(I+1)-P(I))/P(I)
47 LET S(I+1)=S(I)+R
50 NEXT I
55 PRINT " I              P(I)            S(I)"
56 PRINT
57 PRINT
60 FOR I=0 TO 49
70 PRINT I,P(I),S(I)
80 NEXT I
90 END

READY
```

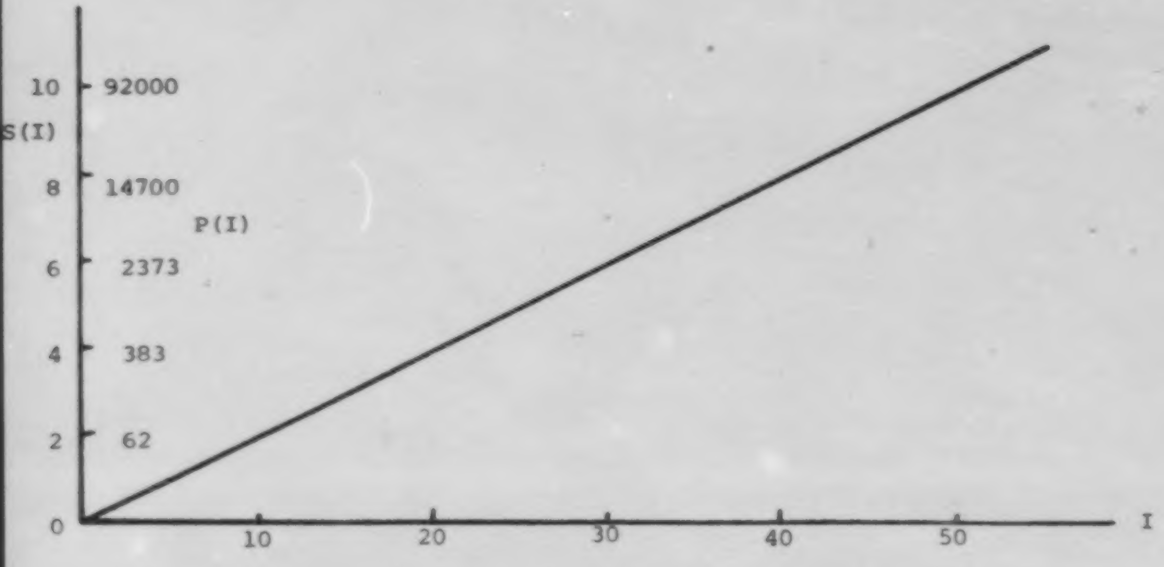


Fig. 1.4

```
1 REM  FINITE RESOURCE MODEL, COMPRESSED SCALE  CALCULATION
10 DIM P(60),S(60),L(60)
20 INPUT G,G1,P(0)
25 LET S(0)=0
30 FOR I=0 TO 49
40 LET P(I+1)=P(I)+(G-G1*P(I))*P(I)
45 LET R=(P(I+1)-P(I))/P(I)
47 LET S(I+1)=S(I)+R
48 LET L(I+1)=LOG(ABS(P(I+1)-P(0)))
50 NEXT I
55 PRINT "   I            P(I)            S(I)            L(I)"
57 PRINT
60 FOR I=0 TO 30
70 PRINT I,P(I),S(I),L(I)
80 NEXT I
90 END

READY
```

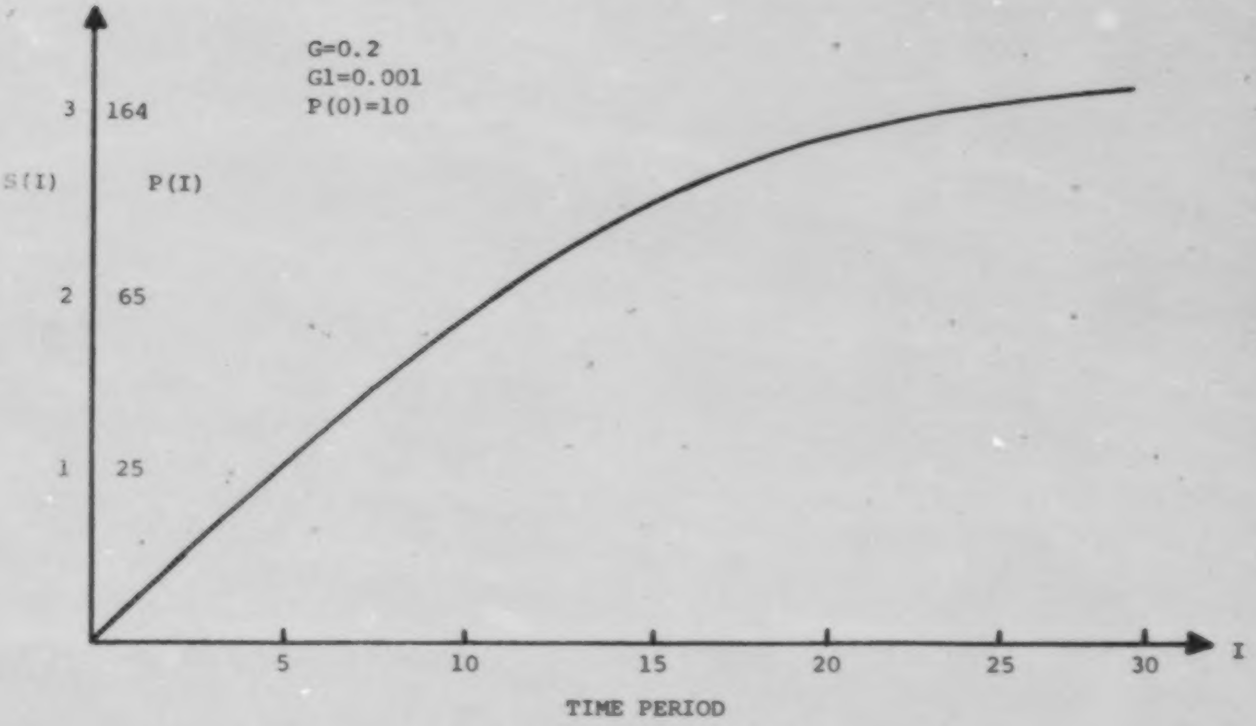


Fig. 1.5a

```
    I            P(I)          S(I)          L(I)

   0             10            0
   1             14.9          .49           1.58924
   2             22.128        .9751         2.49552
   3             32.7023       1.45297       3.12247
   4             47.9841       1.92027       3.63717
   5             69.6736       2.37229       4.08889
   6             99.656        2.80261       4.49598
   7             139.553       3.20296       4.86409
   8             189.854       3.5634        5.19215
   9             248.737       3.87355       5.47536
   10            311.235       4.12481       5.70789
   11            369.985       4.31358       5.88606
   12            418.089       4.44359       6.01148
   13            452.335       4.5255        6.09207
   14            473.896       4.57317       6.13966
   15            486.266       4.59927       6.16598
   16            492.945       4.61301       6.1799
   17            496.423       4.62006       6.18708
   18            498.198       4.62364       6.19072
   19            499.096       4.62544       6.19256
   20            499.547       4.62635       6.19348
   21            499.773       4.6268        6.19394
   22            499.887       4.62703       6.19417
   23            499.943       4.62714       6.19429
   24            499.972       4.6272        6.19435
   25            499.986       4.62722       6.19438
   26            499.993       4.62724       6.19439
   27            499.996       4.62725       6.1944
   28            499.998       4.62725       6.1944
   29            499.999       4.62725       6.1944
   30            500           4.62725       6.1944

READY
```

Results from program listed in Fig. 1.5a

Fig. 1.5b

and initial values of the function. However, this calculation does serve as a very effective and cheap way for compressing the range of a function. This in turn permits an easy graphical analysis of the behavior of the function.

REMARKS CONCERNING THE PROBLEMS AT THE END OF THE CHAPTERS

In working the problems the student is encouraged to use the computer in any manner that he feels will assist him in solving a problem. For the easier problems this suggestion may seem rather wasteful; however, we are not discouraging thinking at the expense of computer use. Rather, we are urging the student to become facile with computer assisted analysis because as the problems become more difficult, the use of simple and straightforward computer methods will be the only practical way of obtaining a solution. A further reason for exhorting the student to use a computer is the fact that the act of formulating a problem for a computer necessitates a thorough understanding of the problem. Plotting or graphing of results is also encouraged and may be done by hand, using graph paper or with the aid of the computer on a teletype, printer, plotter or cathode ray tube.

Some of the problems are quite easy and as many as possible should be attempted. If computer time is limited, the student should define, flowchart and write out the principle BASIC programming language statements for those problems he is unable to solve with the aid of a computer.

Unless otherwise stated, the birth rates, mortality rates and growth rates are given in terms of the "natural" time period suggested by the problem. The time period will usually be a generation, a week and an hour, etc. and should be clearly specified in your work.

PROBLEMS

CHAPTER I

1. Consider an initial population of 100 individuals growing in a constant environment.
   - (a) If the birth rate is 0.8 and the mortality rate is 0.3, how many generations must pass before the population increases 10 times? 100 times?
   - (b) If the population is 150 individuals after a single generation, what is the growth rate?
   - (c) If the population is 75 individuals after a single generation, what is the growth rate?
   - (d) If the population is 300 individuals after 10 generations, what is the growth rate? (HINT: Using the computer program to assist your guessing.)
   - (e) What growth rate will result in a population of 15 individuals after 20 generations?
2. Suppose there are two populations growing in a constant environment with the same growth coefficient and the initial population of the first population is twice the initial population of the second population. How do the number of individuals in each population compare at the fourth generation? If the growth coefficient for each population is 0.1, what is the "doubling time" for each population? The time to increase 10 fold? What do you conclude about the effect on the population growth of different starting populations?
3. In the finite resource model, if the growth rate is 0.8 and the carrying capacity 1000 individuals, what is the auxilliary growth coefficient?
4. Consider an initial population of 200 individuals growing in a finite environment with a growth coefficient of 0.9 and an auxilliary growth coefficient of 0.000045. What is the carrying capacity? How many generations are necessary to achieve the carrying capacity? As the auxilliary growth coefficient is increased, what happens to the carrying capacity? How does the number of generations required to attain carrying capacity vary as the auxilliary growth coefficient increases? (HINT: Make some runs on the computer corresponding to different values for the auxilliary growth coefficients and examine the results.)

5. Consider the finite resource model and let G=0.8, and G1=0.0001. What is the carrying capacity? How does the shape of the growth curve vary as both G and G1 increase in proportion so that the carrying capacity does not change?
6. For the finite resource model, make up an initial population, as well as growth coefficients, and then graph the difference P(I+1-P(I) vs. I. Compare this with a graph of the population. Discuss the comparison.
7. Modify the Malthus model program to permit the alteration of the growth coefficient after every 10 units of time. Denote the growth coefficients by G(K), K=0, 1, 2, 3, 4. Select a set of values for the growth coefficients, G(K).
   (a) Plot P(I) vs. I.
   (b) Plot the sum of the relative changes in the population per period.
8. Modify the Malthus model program in accordance with the following hypotheses.
   (a) The number of births per period is assumed to be proportional to the square of the population.
   (b) The number of deaths per period is assumed to be proportional to the cube of the population.

   Choose various values for the respective constants of proportionality, run the program and discuss the results.
9. With the aid of the Malthus model
   (a) How many generations are necessary for the initial population to increase twenty fold if the growth coefficient is 0.1, 0.5, 1.0?
   (b) How many generations are necessary for the population to become less than one-twentieth of the original population if the growth coefficient is -0.1, -0.5, -1.0?
10. Modify the finite resource model to accord with the assumption that the birth rate will decrease in proportion to the square of the population and the mortality rate will increase in proportion to the cube of the population. Choose the constants of proportionality to be 0.0001 and 0.00001 respectively. Run the program with B=1.0 and M=0.5 and with B=1.0 and M=0.1. Discuss the results.

11. Assuming a time period of one year, modify the finite resource model to include:

    (a) Emigration and Immigration.

    (b) Harvesting and seeding.

    State your hypotheses clearly and indicate how they are implemented in the program. As an example of part a, you might assume that the number emigrating is proportional to the existing population and that emigration would not occur until a certain population has been reached. Similar statements could apply to immigration. As an example of part b, you may want to harvest every 5 years, or every 9 years, and the number you may harvest could be a number chosen at random within certain bounds. It may be helpful to write the fundamental equation as P(I+1)=P(I)=(No. of births)-(No. of deaths)+(No. of Immigrants)-(No. of Emmigrants)+(No. seeded)-(No. harvested) where the quantities in parentheses are measured per period. Now make up your own hypothesis about each of the quantities.

12. In a certain habitat a herd of 1000 grazing animals has been in long term existence. The herd is suddenly transferred to a new grazing area which has a different type of feed grass. It is observed that the weekly change in the herd population is proportional to the amount of new grass in excess of 500 bales. In turn, the weekly change in the amount of new grass decreases in proportion to the number of animals greater than the steady state population of 1000 animals. The initial amount of feed grass available in the new grazing area is 750 bales. With the aid of a computer program and constants of proportionality that you select, describe the time evolution of both the herd population and the amount of new grass - using a time period of one week.

13. A microbial population is growing in a culture and it is observed that the population increases 30% every 3 hours. Using a time period of one hour, what is the population at the end of the 12th hour, the 18th hour, and at the end of the day, if the initial population is 1000?

14. For the first five weeks, an insect colony is observed to increase 25% per week whereas for all succeeding weeks the population decreases 5% each week. Assuming a one week time increment, how many weeks must elapse before the population vanishes if the initial population is 500? 5000?

15. A culture is growing at the rate of 30% per hour. How long will it take for the population to double? To increase by ten fold? By 100 fold? Use one minute time increments.

16. Two bacteria populations, A and B, exist in isolation from each other. Type A bacteria is growing at the rate of 5% per day and type B is growing at the rate of 20% per week. If the respective initial sizes are 10 and 1000 and a time increment of one day is used, assuming both populations start growing on the same day, how long will it be until —

   (a) The sum of the two populations is 100,000?
   (b) Both populations are equal?
   (c) Type B population is twice the size of the type A population?
   (d) Type A population is twice the size of the type B population?

17. The initial population of a group of insects is 1000 and the population at the end of the first generation is 1250.

   (a) If the increase in the population each generation is 50% of the increases in population of the preceding generation, what is the population after 20 generations? After 50 generations?
   (b) Using the hypotheses of part (a), after how many generations will the population be 45 times the initial population?
   (c) Using the hypotheses of part (a) and assuming that there is an emigration of 100 insects each generation, beginning with the 2nd generation, after how many generations will the population be 50 times the original population?

18. A yeast culture is growing in such a manner that, after each hour, the culture increases by an amount equal to 25% of its present size. If the initial culture size is 10 and a time increment of one hour is used,

   (a) How many hours before the culture size is 5000?
   (b) How many hours before the culture size is 25 times the culture size at 15 hours?

19. For I=5 write out the sequence of expressions for T(J) that is generated by the program listed on page 1.17.

REFERENCES

CHAPTER I

D'Ancona, Umberto 1954. The Struggle for Existence, trans by Charles, A. and Withers, R. F. J. E. J. Brill, Leiden, Netherlands.

Allee, W. C. 1938. The Social Life of Animals, W. W. London, New York.

Hazen, William E. 1964. Readings in Population Ecology. W. B. Saunders Co. Philadelphia, Pa.

Grossman, S. I. and Turner, J. E. 1974. Mathematics for the Biological Sciences. MacMillan Co., New York.

Batschelet, E. 1971. Introduction to Mathematics for Life Science Students. Springer-Verlag. Berlin.

Gause, G. F. 1971. The Struggle for Existence. Dover. New York.

Lotka, A. J., 1956. Elements of Mathematical Biology. Dover. New York.

Rashevsky, N. 1960. Mathematical Biophysics. Physics-Mathematical Foundations of Biology. Vol I and II. Dover. New York: 1960.

Story, R. W. and Waxman, B. D. 1965. Computers in Biomedical Research. Vol I. Academic Press. New York.

Poole, R. W. 1974. An Introduction to Quantitative Biology. McGraw-Hill. New York.

McNeary, S. S. 1973. Introduction to Computational Methods for Students of Calculus. Prentice-Hall. Englewood Cliffs, N. J.

Hamming, R. W. 1968. Calculus and the Computer Revolution. Houghton-Mifflin. Boston, Mass.

Don, W. S., Butler, G. D. and Herta, D. L. 1972. Computer Applications for Calculus. Prindle, Weber and Schmidt. Boston, Mass.

Lumbach, L. Carl. 1974. Calculus with the Computer, A Laboratory Manual. Prentice-Hall. Englewood Cliffs, N. J.

CHAPTER II

THE ASSOCIATION OF TWO SPECIES

Independent Growth

The previous chapter considered the development of simple computer models for the simulation of the growth of a single or isolated species. In this chapter, the effect of introducing a second species will be examined. The technique to include the effects of another species will parallel the technique used to develop the models exhibited in the first chapter. In that chapter the simplest assumptions were made and a model devised and a computer program written. The construction of more realistic models was accomplished by systematically removing or modifying some or all of the original simple or restrictive hypotheses. This resulted in a sequence of more and more complicated computer programs. In this chapter, the same procedure will be followed. Consequently, it will be first assumed that such effects as contamination, finite or fixed food supply, etc. are not present. Thus, only the effect of the association of two species in the presence of the same food supply will be considered. Initially, it will be assumed that for all times in the growth period of both species that there is sufficient food for each specie and furthermore, that neither species is a food supply for the other. These hypotheses thus imply that each species will grow independently of the other. This suggests that only simple alterations of the first program should be necessary in order that the program may be used to describe the resultant growth of two species. These alterations are most readily accomplished by introducing the following rather self evident notation.

Let:

P1 denote the population of the first species,

B1 and M1 denote the coefficients of natality and mortality respectively of the first species,

P2, B2 and M2 denote the corresponding variables for the second species.

The student should note that these variables do not have the same meaning as they had in the first chapter.

An examination of the program, page 1.6, reveals that line 40 should be modified to read

```
40  LET P1(I+1)=P1(I)+G1*P1(I)
```

and an additional line, line 45, should be inserted

```
45  LET P2(I+1)=P2(I)+G2*P2(I).
```

In addition, the notation G1=B1-M1 and G2=B2-M2 has been introduced. Of course, lines 20, 25 and 70 must also be altered. The completed program then appears as

```
 1  REM POPULATION GROWTH MODEL.  TWO SPECIES.
10  DIMENSION P1(50), P2(50)
20  INPUT B1, B2, M1, M2, P1(0), P2(0)
25  LET G1=B1-M1:  LET G2=B2-M2
30  FOR I=0 TO 49
40  LET P1(I+1)=P1(I)+G1*P1(I)
45  LET P2(I+1)=P2(I)+G2*P2(I)
50  NEXT I
60  FOR I=0 TO 49
70  PRINT I, P1(I), P2(I)
80  NEXT I
```

Since lines 40 and 45 are uncoupled, i.e. each does not contain any terms present in the other, the results of running this program will be the same as those obtained from the Malthus model. This program together with its development has been presented because it will serve as a basis of developing the programs to be described in the following sections.

Effect of Fixed or Finite Resources

We now wish to consider the case of two populations competing for the same limited resource; for example, food supply.

In order that we may construct a model for the two populations, the previous equations will be modified to include the assumption that the food supply is of limited extent and that neither population is food for the other. The program alterations necessary to accommodate this assumption are very similar to those made when the effect of a finite food supply upon only a single species was considered. The student with a short memory is urged to reread that section.

In order that the student may better understand the derivation of the program alterations, an alternative development of the corresponding alteration of the single specie program will be presented. To arrive at the alteration described on page 1.10, we could have 'argued' or 'reasoned' in the following manner. Certainly, as the population increases, the available food decreases and the magnitude of the decrease in food in one time period should be proportional to the population present in this time period. Thus, the net decrease of food in the I^{th} time period is R*P(I) where R is a constant of proportionality and is positive. This net decrease in food supply in a given time period should reduce the proportion B, of births and increase the proportion M, of deaths in the same time period. This assumes that the time interval in generation periods is sufficiently long so that the change in food supply manifests itself in the same period. The magnitude of the reduction of B will be assumed to be proportional to this decrease in food supply. Thus, the net decrease in B in a given time period may be written as Q*R*P(I) where Q is a new constant of proportionality and is also positive. If the notation B1=Q*R is introduced then the net proportion of births in the time period is given by B-B1*P(I). This is the term that was previously developed. In an analogous manner one may obtain the alteration to the net proportion of deaths. The previous discussion can now be used as a basis for developing the necessary alterations of our program to describe the growth of two species competing for a limited or fixed food supply. The student will note that in the process of developing the necessary alterations to the program, that a new notation will be introduced. He should not be confused by this introduction and should learn to accept such a procedure as a matter of routine. Of course, it is assumed that the notation that is introduced is self explanatory or quite obvious in the meaning it is intended to convey. The integers 1 and 2 following the letters will usually refer to the first and second species respectively.

The amount of food consumed by each species in a time period is

$$R1*P1(I) \qquad \text{and} \qquad R2*P2(I)$$

respectively where $R1$ and $R2$ are constants of proportionality and are positive. The amount of food $T(I)$, consumed by both species in a time period is then

$$T(I)= R1*P1(I)+R2*P2(I)$$

and the assumed net decrease in the proportion of births of the first species in the time period is proportional to this and equal to

$$Q1*T(I).$$

Here $Q1$ is the constant of proportionality relating the decrease in food supply to the decrease in proportion of births. Hence, the proportion of births of the first species in one generation is

$$B1-Q1*T(I).$$

In a similar manner the proportion of births of the second species is given by

$$B2-Q2*T(I)$$

and $Q2$ is a constant of proportionality. Both $Q1$ and $Q2$ are positive and very much smaller than $B1$ and $B2$ respectively.

By using analogous reasoning, it can be seen that the proportion of deaths of each species for the same time period is given by

$$M1+S1*T(I)$$

and

$$M2+S2*T(I)$$

respectively. The constants S1 and S2 are the constants of proportionality relating the change in food supply in the time period to the increase in the proportion of deaths in the time period. They are positive and very much smaller than M1 and M2. If the student "works out on his own" the derivation of the last two expressions, he will assure himself of a good grasp of the result.

With the aid of these results, the necessary alterations to the previous program may now be made to include the effect of a finite environment. Line 40 of page 2.2 should thus be modified to read:

```
40  LET P1(I+1)=P1(I)+(B1-Q1*T(I))*P1(I)-(M1+S1*T(I))*P1(I).
```

If the notation

G1 = B1-M1   and   T1 = Q1+S1

is introduced the previous line may be written as

```
40  LET P1(I+1)=P1(I)+(G1-T1*T(I))*P1(I).
```

The analogous equation expressing the growth law of the second species P2 is:

```
45  LET P2(I+1)=P2(I)+(G2-T2*T(I))*P2(I)
```

where

G2 = B2-M2   and   T2 = Q2+S2.

In addition a new line, call it 38, must be introduced to calculate T(I) the amount by which the food supply decreases in the I^{th} period. Thus,

```
38  LET T(I) = R1*P1(I)+R2*P2(I).
```

Equations 40 and 45 are a pair of coupled equations and hence the task of evaluating their accuracy in describing the growth of the two species by making several runs and examining each run is not trivial and could certainly be time consuming and expensive of computer time. We are thus faced with two problems which will occur again and again in our work. The first is the assessment of the validity of our BASIC equations in describing the phenomena under investigation. The second problem is the checking out or debugging of the computer program. It should be noted that these two problems are usually closely related and difficult to separate. In what follows, we shall try to indicate some procedures and techniques useful for determining the validity of the model and the program. Of course, the principle or most significant assessment of the accuracy of the program as well as the validity of the model is the degree of agreement between the computational results and the empirical data. Consequently most of our discussions concerning validation will be concerned with checking the program. A technique that is frequently used to assist in the validation of the accuracy of the program, (commonly called debugging) consists in examining the basic equations in order to obtain results which can then be used to check the actual numerical results. Such an examination can have several different forms. One of the simplest of these forms consists in specifying certain values of the parameters in order to obtain known results. For example, suppose that in line 40, B1 is set equal to M1 and Q1 is set equal to the negative of S1. Hence G1=0 and T1=0 and so for all I, P1(I+1)=P1(I); i.e. the first population remains constant. Thus, when the constants B1, M1, Q1, and S1 are chosen in the aforementioned manner, and the program is run, the values of the first population should not change. Similar remarks obtain for the appropriate constants and the second population. The student should be able to work out for himself how to choose all of the constants of proportionality in order to insure that both populations change at the same rate, i.e. P1(I)=P2(I) for all I. (This does not mean that the populations remain constant).

We again point out to the student that the simplified form of equations 40 and 45 is not absolutely essential to the operation of the program. The simplification is done to save computational effort and hence computer cost. The simplified form is usually easier to examine and program accurately. On the other hand, it is frequently the case that modifications of the model, and hence of the program, are more readily carried out when the equations appear in their long form. Consequently, when a model is being developed and assessed, the long form is usually the most accessible form to work with. In addition, most of the programs that are developed in this text are short and do not require significant amounts of computer time to execute. Hence for this program, the omission of the simplification of the equations is not so essential. However, for sophisticated and complex models that result in computer programs that are long running and frequently used, such simplification can achieve considerable savings in computational cost.

Now both equations 40 and 45 are simple in appearance; however, these equations still contain a great deal of information which is of assistance in checking the program. To obtain this information we will employ a technique which forms the basis of much of the purely mathematical and theoretical research in the physical and engineering sciences. It is usually always the case in these sciences that the equations describing the phenomena under investigation are too difficult to solve. Consequently, the engineer or scientist must frequently be satisfied with only partial answers to questions whose relevance may be indirect. In attempting to formulate such questions, the physical scientist will often alter the basic equations by rewriting them in different forms and, if possible, then interpret the resulting forms in terms of the original phenomena. This process will sometimes suggest "useful" and "allied" questions whose answers may more readily be obtained. It is in this manner or spirit that we now proceed.

The student will recall that the expressions

```
        P1(I+1)-P1(I)     and    P2(I+1)-P2(I)
```

are the changes in the populations in a given time period; that is, they are the rates of change of the populations. It is perhaps suggestive to examine the expressions for each of these rates of change. The expressions are respectively:

```
        (G1-T1*T(I))*P1(I)
```

and

$$(G2-T2*T(I))*P2(I).$$

Since $T(I) = R1*P1(I)+R2*P2(I)$ and both $R1$ and $R2$ are positive, it is evident that as both populations increase so also will $T(I)$ increase. Furthermore, in the unlikely event that the populations $P1(I)$ and $P2(I)$ were such that both of the equations

$$G1 = T1*T(I) \qquad \text{and} \qquad G2 = T2*T(I)$$

were simultaneously satisfied, the populations would remain constant. An alternative way of expressing this fact is to note that if both of the ratios $G1/T1$ and $G2/T2$ simultaneously equaled $T(I)$ for some value of $P1(I)$ and $P2(I)$ respectively, then the populations would remain unchanged. We have thus shown that there could exist values of $P1(I)$ and $P2(I)$, called critical values, such that the populations would remain constant.

The student should note that it is not the case that if $G1/T1=G2/T2$ that there will necessarily exist populations $P1(I)$ and $P2(I)$ such that $G1=T1*P1(I)$ and simultaneously that $G2=T2*P2(I)$. It is always the case, however, that if both populations remain unchanged that the aforementioned two equations hold.

Problem

Show that if both populations are constant that the two populations are related by

$$\begin{aligned} P2(I) &= \left(\frac{G2}{T2} - R1*P1(I)\right) /R2 \\ &= \left(\frac{G1}{T1} - R1*P1(I)\right) /R2 \end{aligned}$$

Using mathematical techniques from the theory of differential equations Volterra was able to further show that if $G1/T1$ was greater than $G2/T2$ then the second population would approach zero, that is

P2(I) would die out. In this case, he was also able to show that the first population, P1(I), would approach the value G1/(T1*R1).

We now show how an examination of the expressions for the differences in populations at successive time intervals can reveal how Volterra's conclusions may be obtained. These expressions may be written in the alternate forms

T1*(G1/T1 - T(I))*P1(I) (a)

and

T2*(G2/T2 - T(I))*P2(I). (b)

Now, if G1/T1 is greater than G2/T2, then the quantity G1/T1-T(I) is greater than the quantity G2/T2-T(I) and hence the change in the first population is greater than the change in the second population. Since T(I) increases as P1(I) and P2(I) increase, and the quantities G1/T1 and G2/T2 are constant, it is evident that there will be populations P1(I) and P2(I) such that G2/T2-T(I) is negative and that G1/T1-T(I) is positive. For example, suppose that both populations are such that G2/T2>T(I) and G1/T1>T(I). In this event both populations would increase and consequently T(I) would also increase. This increase of both populations would continue until T(I) became greater than G2/T2, at which point the change in the second population would become negative, i.e. P2(I) would begin to decrease. Because G1/T1 is greater than T(I), the first population would continue to increase while the second population continued to decrease until it became zero.

When P2(I)=0, it will remain so and the value of T(I) will be R1*P1(I). It is also evident that when P2(I)=0 there can be no further change in P1(I) since an increase in P1(I) would result in a negative population for the second species. Now, the change in the population P1(I), expression (a) above, can be written in the form

T1*(G1/T1 - R1*P1(I))*P1(I).

Since there is no change in the first population, an examination of this quantity shows that

```
G1/T1 - R1*P1(I) = 0
```

or

```
P1(I) = G1/(T1*R1)
```

which is Volterra's result.

Despite the simplicity of the aforementioned model which describes the association of two species which compete for the same resource (in this case we assumed that the resource was food supply whereas any resource required by both species could have been chosen), the results predicted by the model are verified in nature. Thus, in nature it is rarely found that two closely related species simultaneously exist in the same biotype or locale. This is because one species survives and the other dies out. This is predicted by our model and whichever species lives or dies is dependent upon its natural growth rate and upon the amount of the resource it requires. A possible form of this dependency is given in the form outlined above. A more detailed discussion of species competition is given in Emlen (1973).

One Species Feeds Upon the Other

The previous section investigated the relation existing between two species each of which competed for the same resource. It is natural to attempt to develop a model describing the association between two species wherein one specie subsists upon the other specie. In order to develop such a model, it will be assumed that the food source for the first species is abundant and is not affected by the presence of the second or predator species. It will also be assumed that the second species feeds exclusively upon the first.

In order to obtain some 'feel' or 'intuition' about the construction of an appropriate model, it is instructive to review the development corresponding to two species which are living in isolation from each other according to the Malthus model. The equation describing the growth of the first species is:

```
40 LET P1(I+1)=P1(I)+(B1-M1)*P1(I)
```

where B1 and M1 are the coefficients of natality and mortality respectively for the first species. Since food is abundant for the first species the growth coefficient, G1 = B1 - M1, is positive. A similar equation describes the growth of the second species and is

```
45 LET P2(I+1) = P2(I)+(B2-M2)*P2(I)
```

where B2 and M2 are the corresponding proportion of births and deaths. However, since the species are living in isolation from one another and the first species is the <u>only</u> food supply for the second species, the coefficient of mortality M2 for P2(I), must be much greater than the coefficient of natality B2. Thus, the growth coefficient, G2 = B2 - M2, is negative and the second population will perish. The population curves for P2(I) will resemble those given in figure 1.1b, page 1.7. On the other hand, since the food supply is assumed to be abundant and the environment advantageous for P1, B1 will be greater than M1 and hence, G1 will be positive and the first population will thus increase indefinitely.

We now return to an examination of the interesting case wherein the two species are not in isolation from one another, but rather are existing in the same habitat. This is called the predator-prey problem. Consider the effect of such an association on the first, or prey, population P1. The fact that the first species is the exclusive food supply of the second may be reflected in our model by altering the natality and mortality rate of the first species in proportion to the number of the second species. Hence, it will be assumed that in a given time period, the actual or net increase in the proportion of deaths of the first species will be proportional to the population of the second species and that this increase is given by +D1*P2(I) where D1 is the constant of proportionality. Similarly, the net decrease in proportion of births in a time period will be given by -N1*P2(I) where N1 is the corresponding constant of proportionality. In a similar manner, it is reasonable to assume that in a time period, the changes in the proportion of births and in the proportion of deaths in the second or predator population should be proportional to the number of the first species. Thus, the terms N2*P1(I) and -D2*P1(I) will represent the respective increase in birth rate and decrease in death rate of the second species in a time period. The constants N1, N2, D1 and D2 are all positive and very small compared to G1 and G2. To accommodate these changes, lines 40 and 45 of the previous program are modified to read

```
40 LET P1(I+1)=P1(I)+((B1-N1*P2(I)-(M1+D1*P2(I))))*P1(I)
```

and

```
45 LET P2(I+1)=P2(I)+((B2+N2*P1(I)-(M2-D1*P1(I))))*P2(I)
```

or

```
40 LET P1(I+1)=P1(I)+(G1-(N1+D1)*P2(I))*P1(I)
```

and

```
45 LET P2(I+1)=P2(I)+(G2+(N2+D2)*P1(I))*P2(I)
```

By introducing the notations, F1 = N1+D1 and F2 = N2+D2, these equations are more conveniently written as

```
40  LET P1(I+1) = P1(I)+(G1-F1*P2(I))*P1(I)
```

and

```
45  LET P2(I+1) = P2(I)+(G2+F2*P1(I))*P2(I).
```

It must be recalled that since the prey is the sole available food supply for the predator in the environment, G2 is negative, and that

G1 = B1-M1   and   G2 = B2-M2.

This model is commonly called the Lotka-Volterra model of prey-predator interaction. The next section discusses a second model of prey-predator interaction called the Leslie model which results from imposing the effect of a finite environment upon the prey population.

The listing of the Lotka-Volterra model program is given in figure 2.1. Lines 42 and 47 have been inserted to halt the calculation whenever either population becomes extinct. The heart of the program is contained in statements 35 to 60. The remaining statements provide for storage, input, printing and output.

Since the quantity F1*P2(I) increases or decreases as P2(I) increases or decreases, the constant F1 is a measure of the effectiveness with which the first species defends itself from the predator species P2(I), and hence F1 is called the defense coefficient. If the defense coefficient is very small, the predator population must become quite large before the net growth rate of the prey is significantly decreased. Conversely, if the defense coefficient is large, a small increase in the predator population will result in a large decrease in the net growth rate. For analogous reasons, F2 is called the voracity or offense coefficient. These coefficients are such that if the prey, P1(I) improves his defensive mechanism both

```
1 REM    TWO SPECIES PREY-PREDATOR MODEL
10 DIM P1(100),P2(106)
20 PRINT "TYPE  G1,  G2,  F1,  F2,  P1(0),  P2(0)"
21 PRINT "REMEMBER, G2 MUST BE NEGATIVE"
22 PRINT
24 INPUT G1,G2,F1,F2,P1(0),P2(0)
25 PRINT
30 PRINT "TYPE NO. OF TIME PERIODS TO RUN"
32 PRINT
34 INPUT N
35 FOR I=0 TO N
40 LET P1(I+1)=P1(I)+(G1-F1*P2(I))*P1(I)
42 IF P1(I+1)<0GO TO 90
45 LET P2(I+1)=P2(I)+(G2+F2*P1(I))*P2(I)
47 IF P2(I+1)<0GO TO 100
60 NEXT I
62 PRINT
63 PRINT
65 PRINT "TIME PERIOD, PREY POP., PREDATOR POP."
66 PRINT
70 FOR I=0 TO N
75 PRINT I,P1(I),P2(I)
80 NEXT I
85 GO TO 200
90 PRINT "THE PREY POPULATION BECAME NEGATIVE"
92 GO TO 200
100 PRINT "THE PREDATOR POPULATION BECAME NEGATIVE"
105 GO TO 200
200 END

READY
```

Lotka-Volterra Model

Figure 2.1

coefficients will decrease whereas if the predator, P2(I) improves his predatory capability both coefficients will increase. In order to obtain a quantitative feeling for these statements, suppose that the magnitude of the growth coefficient G1 is 0.5, the magnitude of F1 is 0.001 and that the predator population P2(I) is 100. Hence, the net or modified growth rate, G1-F1*P2(I), is 0.4. Now if the predator population is doubled, the modified growth rate becomes 0.3 and the net growth rate is decreased by 25%. If, however, the magnitude of F1 is 0.002, then the net growth rates corresponding to predator populations of 100 and 200 respectively are 0.3 and 0.1. Thus, doubling the defense coefficient results in a reduction of 67% in the net growth rate when the population of the predator is doubled.

Problem

Carry out a similar numerical analysis for the voracity coefficient.

The student is urged to make several runs with the prey-predator program each time using a different set of parameter values. The time history of both P1(I) and P2(I) should be recorded. It will be seen that the predator as well as the prey population increases and then decreases and that this behavior is repeated. Volterra, by using sophisticated mathematical techniques, was able to give a rather complete qualitative discussion of this oscillatory or periodic motion. There is a very extensive literature devoted to the mathematical analysis of such motions.

As in the previous section, it is helpful to analyze the governing equations to ascertain possible or probable results that may be of assistance in assessing the validity of the computer results. This will be done by examining the respective changes in the populations in a given time period. They are

(G1-F1*P2(I))*P1(I)

and

(-G3+F2*P1(I))*P2(I)

where for convenience in the analysis, the notational substitution G2 = -G3 has been made. This change in notation implies that G3 is positive and serves to emphasize the fact that in the absence of the prey population the growth coefficient of the predator population is negative. An examination of these two quantities reveals that if P1(I) and P2(I) were ever such that

$$P2(I) = G1/F1$$

and simultaneously

$$P1(I) = G3/F2,$$

then the populations would remain unchanged since there would be no changes in either the prey or the predator populations. Furthermore, if the respective changes in the populations in a given time period are written in the form

$$F1*(G1/F1-P2(I))*P1(I)$$

and

$$F2*(-G3/F2+P1(I))*P2(I)$$

it is easier to perceive why periodic motion is to be expected.

These expressions can also be used to show that neither change in population can indefinitely increase nor indefinitely decrease. Such behavior can be ascertained by noting that if, for example, the prey population were to increase indefinitely, the quantity

$$-G3/F2+P1(I)$$

would eventually become positive and hence the predator population would begin to increase. A continual increase in the predator population would result in the quantity

$$G1/F1-P2(I)$$

becoming negative. Since G1/F1-P2(I) is directly proportional to the change in the prey population in a time period, the change in the prey population would then become negative; thus causing the prey population to decrease. A similar discussion can be made to show the impossibility of an indefinite decrease of either population. Therefore, the populations must oscillate as time increases. This discussion is quite heuristic in nature and most certainly does not constitute a rigorous proof. However, the discussion does show how a simple analysis can yield useful qualitative results. The computer program should confirm such behavior.

In the following assignment, the student is not expected to derive or obtain BASIC programming language expressions for the answers to the questions. If he can obtain such expressions of if he is familiar enough with mathematics to do so, so much the better. The student is being asked to alter the pertinent constants, make the necessary or sufficient number of computer runs with the altered constants, and to then examine the numerical results to determine the magnitude of the change in the quantity of interest resulting from a prescribed change in magnitude of a particular constant. By doing this, it is intended that the student gain insight and intuition about the model and the phenomena. The determination of the change in one variable, or a set of variables, due to a change in another variable is a very fruitful way to obtain insight. Even though it may require several computer runs and much computing, the tremendous calculational speed of the computer usually enables such a determination to be readily made.

Programming Assignment

Using the prey-predator model

1. (a) Choose an appropriate set of constants and initial conditions and construct and run the program.

(b) Graphically display the results to confirm the conclusions stated above.

(c) Vary some of the parameters and/or initial populations and discuss the differences in the results.

⊛ (d) The period is the length of time necessary for the populations to repeat themselves.
By varying some or all of the constants, can you determine how the period depends upon the parameters?

⊛ (e) How do the maximum and minimum populations depend upon the parameters?

2. (a) Modify your program to include the ability to harvest either the prey or the predator.

(b) State the hypothesis used in obtaining the expressions for the harvesting.

(c) By varying the amount of harvest, can you control the population of the prey, the predator, or both, in a prescribed way? This is a problem which concerns wildlife or game managers.

3. (a) Let the magnitude of the prey population be the abscissa and the magnitude of the predator population be the ordinate and plot the time evolution of the populations. What do you observe? Such a plot is frequently called a phase plane plot and is very helpful in analyzing periodic or near periodic motions.

⊛ indicates difficult problems

A Leslie Type Model

By altering the hypotheses, Leslie obtained a different set of equations describing the interaction of a prey and a predator population. The Leslie model is derived by first assuming that the prey population exists in a finite environment and that, therefore, the natural growth of the prey must be moderated by the limited resources of the finite environment. Thus, the fundamental growth equation for the prey, in the absence of a predator population, is

```
40  P1(I+1) = P(I)+(G1-G3*P1(I))*P1(I)
```

where G1 and G3 are the natural and the auxiliary growth coefficients, respectively. The effect of the predator on the prey population was imposed by Leslie in a manner analogous to that used by Lotka and Volterra. Hence, the fundamental equation for the prey population is altered to read

```
40  LET P1(I+1) = P1(I)+(G1-G3*P1(I)-F1*P2(1))*P1(I).
```

Here F1 is a constant of proportionality relating the number of predators to the proportion of prey taken by the predators in a time interval. This equation is identical in form to the equation labeled line 40 on page 2.5 To verify this assertion, the student should substitute line 38 on page 2.5 into line 40 on page 2.5 and rearrange the terms. The fact that these equations have the same form is a result of the extreme generality and lack of specificity of the analysis. Most quantitative models of bio-science phenomena suffer from being gross generalizations of what is actually occurring. The proper integration of all the interacting phenomena, even assuming that the interactions were known, is a near insurmountable task. Consequently, extreme simplifications and generalizations are made and this can result in the derivation of the same form of a model from two divergent sets of assumptions.

To account for the effect of the prey population on the growth of the predator population, Leslie proceeded in the following way. He reasoned that if there were many predators for each prey, i.e. the ratio P2(I)/P1(I) was large, then the scarcity of food would result in a decreased birth rate and an increased mortality rate for the predator population. Conversely, if P2(I)/P1(I) was much less than one, the consequent abundance of prey relative to the number of predators, should result in an increase in the birth rate and a decrease in the mortality rate of the predator population.

The simplest alteration of the predator birth rate to accommodate the preceding assumption is obtained by subtracting from the natural birth rate B2, a term which is proportional to the ratio of the two populations. Thus, the birth rate of the predator may be written as

B2 - R1*P2(I)/P1(I).

Similarly, the natural mortality rate, M2, is altered to read

M2+R2*P2(I)/P1(I).

In these two expressions, R1 and R2 are the constants of proportionality. In terms of these new birth and mortality rates, line 45 may now be written as

```
45 LET P2(I+1)=P2(I)+(B2-R1*P2(I)/P1(I))*P2(I)
                -(M2+R2*P2(I)/P1(I))*P2(I).
```

With the aid of the notation

R3 = R1 + R2

this equation may finally be written as

```
45  LET P2(I+1)=P2(I)+(G2-R3*P2(I)/P1(I))*P2(I).
```

The alteration of the program shown in figure 2.1 to accommodate the Leslie form of the equations is readily accomplished since only lines 40 and 45 need to be changed in the manner indicated immediately above. Provision must also be made to enter the constants G3 and R3 into the program.

The behavior of the populations as described by the Lotka-Volterra model was characterized by a sequence of oscillations whose amplitudes increased in time. In contrast, both populations in the Leslie model tend to reach an equilibrium after a series of damped oscillations. Such behavior more nearly approximates prey and predator populations living in consort with one another.

Effect of Emigration

There are several possible causes of emigration from a population and an accounting of all of these causes together with their induced emigrations is a difficult task. Intuitively, it seems reasonable to assume that a principle cause of emigration is crowding and that the degree of crowding should directly affect the rate of emigration. The degree of crowding, or population density, is measured in the number of animals per unit area. For ease of treatment, it will be assumed that the area in which the competing populations exist is constant, that is, it does not change from time period to time period. It is convenient to take the magnitude of this area as unity since the magnitudes of the prey and predator populations and the respective population densities are then numerically equal. Because it is reasonable to assume that increasing crowding results in increasing emigration it will be assumed that the number of animals leaving a population in any time period will be proportional to the number of animals in the population during that time period. Hence, the number of emigrants of the prey and predator populations may be denoted by L1*P1(I) and L2*P2(I) respectively, where L1 and L2 are constants of proportionality. The accounting of emigration is then readily included in the program by altering lines 40 and 45, page 2.12 to read

```
40  LET P1(I+1) = P1(I)+(G1-F1*P2(I))*P1(I)-L1*P1(I)
```

and

```
45  LET P2(I+1) = P2(I)+(G2-F2*P1(I))*P2(I)-L2*P2(I).
```

These equations may be simplified by introducing the notation G4=G1-L1 and G5=G2-L2 and recombining terms to give

```
40  LET P1(I+1) = P1(I)+(G4-F1*P2(I))*P1(I)
```

and

```
45  LET P2(I+1) = P2(I)+(G5-F2*P1(I))*P2(I).
```

A comparison of these equations with those given on page 2.12, reveals that the structure of the sets of equations is identical and thus the two models are actually the same and differ only in the magnitude of the growth coefficients.

86

The emigration model that we have developed assumes a continual emigration no matter how dense or sparse the actual population. It is more realistic to assume that there exist prey and predator populations below which there will not be an emigration but rather there will exist an immigration, i.e. an influx of prey or predators. The magnitudes of these critical populations would have to be obtained from experiment or observation. We will denote them by P3 and P4 respectively.

Now the number of animals entering or leaving the area in a time period should be proportional to the magnitude of the difference between the present population and the critical population. Let L3 and L4 be the constants of proportionality for the prey and predator populations respectively. Then, in a single time period, the number of prey entering or leaving the area is

```
-L3*(P1(I) - P3)
```

and the number of predators emigrating or immigrating is

```
-L4*(P2(I) - P4).
```

The fundamental equations become

```
40  LET P1(I+1) = P1(I)+(G1-F1*P2(I))*P1(I)-L3*(P1(I)-P3)
```

and

```
45  LET P2(I+1) = P2(I)+(G2-F2*P1(I))*P2(I)-L4*(P2(I)-P4).
```

These equations may be simplified to

```
40  LET P1(I+1) = P1(I)+(G4-F1*P2(I))*P1(I)+L5
```

and

```
45  LET P2(I+1) = P2(I)+(G5-F2*P1(I))*P2(I)+L6
```

where L5 = L3*P3 and L6 = L4*P4.

We again remind the student that the growth coefficient for the predator, G2, is negative.

Programming Assignment

1. (a) Write or modify a previous program to include the effect of immigration or emigration.
   (b) Choose an appropriate set of parameters and initial values and run the program. Compare the results with those obtained from the prey-predator model.
   (c) Do there exist any pre and/or predator populations which remain unchanged?
   (d) Discuss how the results obtained in part (b) may have been anticipated by an analysis of the fundamental equations. (HINT: The discussion should be similar to that given on pages 2.14 and 2.15 for the prey-predator case).

Environmental Toxicity

The effect of the catabolic agents produced by a population growing in a confined environment with a limited resource was considered in the first chapter. A review of that work should prove advantageous in understanding the development which is to follow. In that work, it was assumed that the amount of contaminant created in a time period was proportional to the existing population in that period. This assumption will also be made in the multipopulation model and thus the amounts of contaminant created in the J^{th} period by the first and second populations respectively, are S1*P1(J) and S2*P2(J) where $0 \leq J \leq I$ and S1 and S2 are the corresponding constants of proportionality.

Just as in the single population contamination model, the effect of the toxicity created by both the first and second populations will be to decrease the birth rate and to increase the mortality rate of each population. The total toxicity at the beginning of the I^{th} period, created by each population, will be denoted by T1(I) and T2(I) respectively. The toxicity of the amount of contaminant created by the first population during the J^{th} time period will be assumed to be D1(I-J)*S1*P1(J). Similarly, D2(I-J)*S2*P2(J) will represent the toxicity created by the amount of contamination deposited by the second population in the J^{th} time period. Thus, the total toxicities are given by

```
120  LET T1(I) =  0
130  LET T2(I) =  0
140  FOR J = 0 TO I-1
150  LET T1(I)    = T1(I)+D1(I-J)*S1*P1(J)
160  LET T2(I)    = T2(I)+D2(I-J)*S2*P2(J)
170  NEXT J
```

The statement numbers only serve to indicate the order of the statements and do not refer to their place in any particular program.

The effect of T1(I) and T2(I) is to adversely alter the birth and mortality rates. In order to more easily account for this effect, it is convenient to introduce the following notation. Let B(K,L) and M(K,L) denote the magnitude of the modification of the birth rate and the mortality rate respectively of the K^{th} species due to the cumulative toxicity of the contaminant deposited by the L^{th} species. Thus, we can write

```
B(1,1)=Y1*T1(I)          B(2,1)=Y3*T1(I)
B(1,2)=Y2*T2(I)          B(2,2)=Y4*T2(I)
```

where Y1, Y2, Y3 and Y4 are constants of proportionality relating the deleterious effect of the contaminant toxicity to the birth rate. Similar equations can be written for the adverse effect on the mortality rate. They are

```
M(1,1)=Z1*T1(I)          M(2,1)=Z3*T1(I)
M(1,2)=Z2*T2(I)          M(2,2)=Z4*T2(I)
```

where Z1, Z2, Z3 and Z4 are the constants of proportionality. These modifications in the birth and mortality rates of each species are incorporated into the fundamental equations in a manner entirely analogous to that used in including the contamination effect in a single population. For ease of writing the equations, we do not include emigration nor immigration nor do we include the effects of a finite food supply. The fundamental equations for each population now become

```
40  LET P1(I+1)=P1(I)+(G1-F1*P2(I))*P1(I)-(B(1,1)+B(1,2)+M(1,1)+
     M(1,2))*P1(I)
```

and

```
45  LET P2(I+1)=P2(I)+(G2+F2*P1(I))*P2(I)-(B(2,1)+B(2,2)+M(2,1)+
     M(2,2))*P2(I).
```

It is important to note that in these equations, the variables B(K,L) and M(K,L) are sums and depend on I, the present time period, and thus are different for every time period. Consequently, the sums must be recalculated every time period and hence it is seen that the retarded time effect manifests itself in a considerable increase in computing effort. If however, the effect of the contaminant is independent of its age or time of existence, then the time delay variables D1(I-J) and D2(I-J) are constant. In this event, the sums do not have to be completely recalculated each period; they merely have to be augmented by the addition of a single term. This term is the amount of contamination created in the present time period.

The most difficult part of developing a useful computer program is the obtaining of the time delay parameters and the constants of proportionality. Each of these parameters is a function of the retarded time (I-J), and the determination of this functional relation or dependence is a difficult experimental task.

We do not continue the discussion of the computer model for the effect of environmental toxicity on the multipopulation because of the aforementioned difficulty of obtaining realistic data. The preceding development is an example of the fact that it is quite easy to construct a very elaborate hypothesis and to then develop a computer program whose usefulness is questionable or difficult. This observation is made to impress upon the student that the theoretician or modeler who constructs the computer program must be aware of the limitations imposed upon the experimentalist in his efforts to obtain the necessary parameters and experimental results. The fact that computer programs based upon elaborate hypothesis may be readily constructed is not without merit however. It is frequently the case that not all of the important variables can be measured experimentally because of excessive cost, excessive time or inability to actually obtain by direct means all of the desired data. Thus, it is necessary to use computer models together with available experimental methods to carry out the investigation. This is almost always the case in the physical and engineering sciences and will most certainly be the case in the life sciences. As quantitative understanding becomes deeper, such understanding will indicate the direction and kinds of experiments to be performed. For these reasons, it is also necessary for the experimentalist to have a working knowledge of the capability and limitations of the digital computer. These facts again illustrate the principle that the experimentalist and the theoretician must work closely together and the more knowledgeable each is of the other's work, the better will be the scientific work.

Other Models

There are many other models of the interaction of two or more populations. For a rather complete discussion of the interaction of two populations, see the work of Murdoch and Daten (1975). In this section we consider some of these models. The first model is an alteration of the prey-predator model to include the assumption that the prey population has a refuge. Thus, it is assumed that a certain number of the prey are always safe from the predator. In order that the student may more easily follow the development, we present an alternative method of deriving the Lotka-Volterra equations. It will be recalled that the prey and the predator equations were first written in the Malthus form and then modified in accord with the assumptions given on pages 2.11 and 2.12. The Malthus form of the prey equations is

$$P1(I+1)=P(I)+G1*P1(I).$$

We now reason in the following manner. The term $G1*P1(I)$ represents the change in the prey population in one time period and it is this term which is to be modified to account for the presence and actions of the predator population. It "seems" reasonable to assume that the change in the prey population should be reduced in proportion to the number of possible interactions of the prey with the predator. The number of possible interactions is $P1(I)*P2(I)$. It is known that the actual number of such interactions resulting in the death of a prey is only a small fraction of the total number of interactions or else the prey population would become extinct. If $F1$ denotes the fraction, the number of actual deaths in a period is then $F1*P1(I)*P2(I)$. Thus, the change in the prey population in a time period is

$$G1*P1(I)-F1*P1(I)*P2(I).$$

If this term is substituted for the change in the prey population as given by the Malthus form of the equation, the equation governing the prey population may be written as it appears in line 40 of the program listed in Figure 2.1. The equation governing the time evolution of the predator may be derived in a similar manner.

We now return to the problem of deriving equations for the prey-predator interaction assuming that the prey has a refuge. Let K denote the number of prey which are assumed to have a safe place for refuge; then (P1(I)-K) denotes the number of prey available for the predator. Thus, the number of possible interactions between the prey and the predator is (P1(I)-K)*P2(I). Since only a fraction, F1, of these interactions actually occur, and we are assuming that each of these results in the death or non-birth of a prey, the change in the prey population in one time period is given by F1*(P1(I)-K)*P2(I). By analogous reasoning, the change in the predator population due to the presence of the prey population, is given by F2*(P1(I)-K)*P2(I) where F2 is a constant of proportionality. Since these terms represent the changes in the prey and the predator populations respectively due to interaction affects, they must each be added to the corresponding Malthus form of the prey and predator populations. Thus, lines 40 and 45 are modified to read

```
40  LET P1(I+1)=P1(I)+G1*P1(I)-F1*(P1(I)-K)*P2(I)
```

and

```
45  LET P2(I+1)=P2(I)+G2*P2(I)+F2*(P1(I)-K)*P2(I).
```

We remind the students that the respective magnitudes of F1 and F2 in the two models are not the same and that G2<0 since it is assumed the predator cannot survive in the absence of the prey. Provision must also be made in the program for inputting K, the number of prey which can have a safe refuge. This alteration, and the replacement of lines 40 and 45 as indicated above, are the only modifications required of the program listed in figure 2.1.

It is also possible to construct a model in which a fraction of the prey population is able to find refuge from the predator population. Before doing so, it will be easier for the student to follow the derivation if we first present derivations of alternative models for the constant environment model, the finite resource model and the prey-predator model. The fundamental equation governing the growth of a population in a constant environment was

$$P(I+1) = P(I)+G*P(I)$$

where $G=B-M$ and where B and M were the respective numbers of births and deaths per individual per time period. The above equation may be written as

$$P(I+1) = (1+G)*P(I)$$

or

$$P(I+1) = R*P(I) \tag{1}$$

where $R=1+G$. The constant R may be conveniently thought of as a population multiplier. If $R>1$, which corresponds to $G>0$ or $B>M$, the population increases as I increases, and if $R<1$, which corresponds to $G<0$ or $B<M$, the growth decreases. Furthermore, since $M\leq 1$ and $B\geq 0$, we must have $R\geq 0$. Equation (1) is the desired alternative form of the Malthus model and is mathematically equivalent to the form given by equation (1) in the first chapter.

To derive, with the aid of equation (1), an equation describing the growth of a population in a finite environment, we recall the derivation given in the previous chapter. In that derivation, the birth and the death rates appearing in the Malthus model were decreased and increased respectively in proportion to the size of the population. This, in effect, decreased the growth rate G, in proportion to the population size and enabled the growth rate to be written as (G-G1*P(I)). Our alternative derivation of a model for population growth in a finite environment, will be based upon the idea of altering the population multiplier R appearing in equation (1) in accord with the consequences of the assumption of a finite resource. Thus, we want the multiplier to decrease as P(I) increases. Such a variation may be accomplished by dividing R by a term which increases as the population increases. We choose the form of the term to be (1+A*P(I)) where A>0 and thus the multiplier may be written as

$$R/(1+A*P(I)).$$

The magnitude of A is very much less than the magnitude of R. The student will note that as P(I) increases, the value of the total fraction will decrease. Equation (1) may then be written as

$$P(I+1)=R*P(I)/(1+A*P(I)). \qquad (2)$$

This is the desired alternative form of a model for the growth of a population in a finite environment. It is important that the student note that this model is actually mathematically different from the previous finite resource model as given by equation 40, page 1.10, since the equations do not give the same numerical results. It is also important to note that dividing R by the term (1+A*P(I)) is just one of many possible ways to get the multiplier to decrease as the population increases. For example, we could have chosen to replace R by the term R/(1+A*P(I)↑2) or by the term R/(1+A*P(I))↑2, etc. The primary reason for replacing R by the term R/(1+A*P(I)) rather than some other term is that this term is very simple. This is in keeping with the philosophy that in the absence of other knowledge or evidence the simplest model is to be preferred.

The condition for the population to "level off" is equivalent to stating that there be no change in the population from one time period to the next. This is obtained by imposing the equality,

$$P(I+1) = P(I)$$

in equation (2). We then get

$$P(I) = R*P(I)/(1+A*P(I))$$

or after some simple algebra,

$$P(I) = (R-1)/A.$$

The value of I for which this holds is the number of the time interval for which there is no longer any change in the population. The population corresponding to this condition is called the carrying capacity for this model. Thus, the carrying capacity is given by (R-1)/A.

An alternative model for the competition of two populations for the same limited resources may be obtained in an analogous manner. We begin the derivation by writing down the equations describing Malthus type growth for each of the populations. The equations are

$$P1(I+1) = R1*P1(I) \tag{3}$$

and

$$P2(I+1) = R2*P2(I).$$

The restrictions on the respective magnitudes of R1 and R2 are analogous to those given previously for R. The imposition of the assumption that both populations are competing for the same limited resources may be accomplished by dividing each of the population multipliers by a term which increases as each of the populations increase. One of many such terms is

```
1 + A1*P1(I) + B1*P2(I).
```

The two preceding equations may then be written as

```
P1(I+1) = R1*P1(I)/(1+A1*P1(I)+B1*P2(I))          (5)
```

and

```
P2(I+1) = R2*P2(I)/(1+A2*P1(I)+B2*P2(I))          (6)
```

Both A1 and A2 are positive and very much smaller in magnitude than R1 and R2 respectively. A1 and A2 are analogous to the logistic parameter A in the single population logistic model. The parameters B1 and B2 are also positive and very much smaller in magnitude than R1 and R2 respectively. They account for the variation in growth of each population due to the magnitude of both populations and are not to be confused with their definitions as given in previous sections. Equations (5) and (6) are alternative forms of the equations describing the growth of two populations competing for the same limited resources.

These equations may be modified in several ways to simulate the interaction of a prey and a predator population. We illustrate one possible modification which leads to the Leslie model of prey-predator interaction. Leslie (1948) chose not to modify equation (5), the equation describing the competitive effect of the predator on the growth of the prey population. However, he did significantly modify the equation describing the competitive effect of the prey on the predator, equation (6). He assumed that a significant factor modifying the growth of the predator population was the presence of the prey as a necessary food supply for the predator. He further assumed that the effect on the growth of the predator of the presence of the prey should be such that if there were a large number of the prey as compared to the number of predators, that the growth of the predator population would increase. Conversely, if the ratio of the number of predators to the number of prey was large, the predator population should then decrease. He accomplished such a variation by rewriting equation (6) as

$$P2(I+1) = R2*P2(I)/(1+C1*P2(I)/P1(I)) \tag{7}$$

where C1 is a positive constant. Equations (5) and (7) are identical with the equations given by Leslie in 1948 and 1958.

An examination of equation (7) shows that if the prey become very numerous compared to the predators, the term $P2(I)/P1(I) \to 0$, and the predator population is then governed by the equation

$$P2(I+1) = R2*P2(I).$$

Analogously, if the prey becomes scarce, or disapperars, the quantity $P2(I)/P1(I) \to \infty$ and equation (7) reduces to

$$P2(I+1) = 0.$$

In this event, the predator population becomes extinct.

We are now ready to derive a model for the interaction of a prey and predator population assuming that a constant fraction of the prey have a refuge from the predator. A convenient starting point is Leslie's prey-predator model. We assume that a fraction of the prey can find a refuge from the predator and hence are not competing with the predator for the limited resources. This fraction will be denoted by (1-K), where $0<K<1$. Thus, only the fraction K of the prey population is subject to predation and competition and the growth of the prey in refuge is similar to that of a single population living in a finite environment. These assumptions suggest that the governing prey equation be written as

```
          P1(I+1)=(1-K)*R1*P1(I)/(1+A1*P1(I))
                  + K*R1*P1(I)/(1+A1*P1(I)+B1*P2(I))        (8)
```

The student will note that the first term on the right hand side of the equation corresponds to that part of the prey population living in the refuge and the second term corresponds to the portion of the prey population competing with the predator population for the limited resources. Equations (7) and (8) are the desired equations and are analogous to those appearing in Leslie and Gower (1960). By varying K it is possible to study the effect of different refuge percentages on the population growth.

Many other modifications to these models can be made. The student is urged to invent some of his own modifications. The establishment of which model is the "best" is a question which your author will dodge since each of these models grossly simplifies the actual state of affairs. In addition, the meaning of the word "best" in this context is quite ambiguous since the word can have many meanings depending upon the circumstances.

We purposely did not indicate how to modify any of the existing programs to include the above models since such modifications require only very simple programming changes. We again remind the student that the obtaining of the constants appearing in each of these models may indeed be difficult. For more about this problem, see the end of this chapter.

Running of Programs

The difficulty of obtaining realistic parameter values and/or initial conditions (starting populations) frequently prevents the ready running of programs such as these. This is due to the fact that the programs usually produce biologically meaningful results only for a restricted range of parameter values and/or initial conditions. For example, in the Lotka-Volterra prey-predator model, it is possible to choose the magnitudes of the voracity and the defense coefficients so that both the prey and the predator populations will oscillate and during part of the oscillations both populations will be negative. Such a result is biologically impossible and demonstrates that the validity of the model depends upon the magnitude of the parameters as well as on the fundamental equations. It is of interest to point out that in a traditional mathematical development of a quantitative methods course, the emphasis is on obtaining a "closed form" expression for the answer. This form will contain the parameters and the intial conditions and because of the difficulty of performing the tedious arithmetic necessary to evaluate the solutions, it is rarely the case that numerical estimates of the answers are obtained and examined. Thus, the problem of obtaining realistic values for the model parameters is really omitting. With the advent of the computer, the task of performing the tedious arithmetic has been eliminated and mathematically oriented quantitative methods courses are placing more and more emphasis on discussing and comparing numerical results. As a consequence, the problem of the determination of parameters which give biologically significant results is assuming greater importance.

To illustrate the preceding discussion, we consider the selection of a set of values for the constants of proportionality occurring in the prey-predator model as listed in fig. 2.1. An examination of equations 40 and 45 shows that it is the growth coefficients G1 and G2, and the modified growth coefficients F1 and F2, that are important. This is equivalent to saying that the quantities of interest are the differences of the respective birth and death rates and the respective sums of the pairs of constants of proportionality, N1, D1 and N2, D2.

The determination of G1 is easy, since, by hypothesis, it is the normal growth rate of the prey in the absence of the predator. G2 is negative because it is assumed that the prey is the only food for the predator. In the absence of experimental data, it is difficult to estimate a reasonable magnitude for G2. Thus, as a first guess, we arbitrarily select a value approximately equal to the negative of the normal growth rate of the predator assuming that its food supply was plentiful. Values for F1 and F2 are then obtained by trial and error. To assist in their determination, we recall that for the finite resource model the normal growth coefficient was very much larger than the modified growth coefficient.

Consequently, it seems reasonable to initially select values for the defense and voracity coefficients that are much smaller in magnitude than G1 and G2 respectively. By examining the results, these values can be improved by iteration. Chapters III, IV and V illustrate other methods for determining model parameters. There are also sophisticated experimental techniques which can be of assistance in the determination of F1 and F2.

By observing the interaction of prey-predator populations, it has been noted that the magnitude of both populations oscillate. By utilizing strict experimental controls, it is possible to achieve a nearly constant environment in which the maximum and the minimum populations are clearly discernable. However, in the real world the environment is anything but static and controlled, and thus the maximum and minimum populations may vary considerably. These observations suggest that a possible aid to the selection of a set of parameter values is the existence of oscillatory-like behavior of each of the populations. The existence of oscillatory-like behavior implies the possibility of the existence of a set of parameter values for which the amplitude of the oscillation remains constant. In an attempt to see if indeed such constant amplitude oscillations can be produced by the model, the student may try several different sets of parameter values. Such an attempt will not be successful. However, it is possible that, due to repeated failure to obtain constant amplitude oscillations, the student will gradually get the idea that it may very well be impossible to find such a set of parameter values. The student who carries out this search and by so doing concludes that it is not possible for the model to produce oscillatory-like behavior, illustrates the use of a model to gain understanding. The lack of existence of constant amplitude oscillations is due to the fact that the governing equations are difference equations and it is known that solutions to this particular set of difference equations do indeed have the property that they oscillate with increasing amplitude. In this regard, see the paper by Innis. In contrast, if the Lotka-Volterra prey-predator problem is formulated in terms of differential equations, it is possible to choose the constants of proportionality so that the amplitudes of the solutions do remain bounded. Thus, in this latter formulation, the prey and the predator populations are repeated and the motion is cyclic.

The prey-predator interaction is best revealed with the aid of a graph whose abscissa and ordinate are respectively the magnitudes of the prey and the predator populations. Figure 2.2 shows such a representation for a typical prey-predator problem.

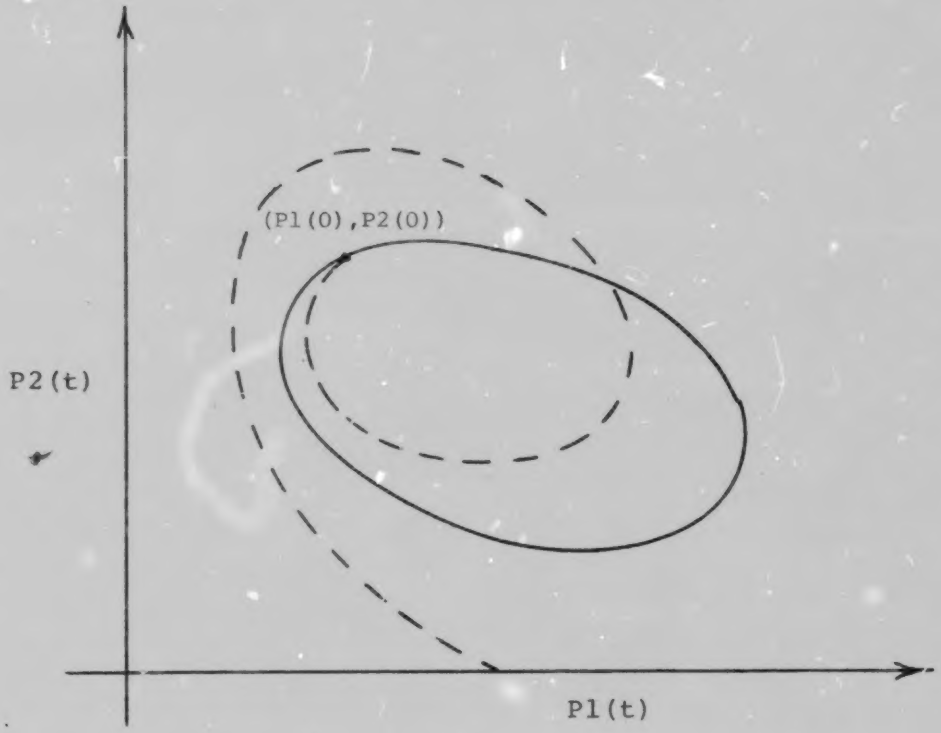


Fig. 2.2

The solid line represents results obtained from a differential equation formulation of the problem using a given set of parameter values. The dashed line indicates results obtained from the BASIC programming language formulation of the problem. In both examples the initial populations were the same. In this regard note part 3a of the programming assignment.

Closing Comments

In closing this chapter, we again reiterate some of the fundamental philosophy concerning the value of models. In essence, we are hypothesizing specific interactions or relations among the populations and their environment and then using the computer to analyze the consequences of such hypotheses. The hypotheses are expressed in the BASIC programming language and are usually very much oversimplified and restrictive. Nevertheless, we feel that they have in them some of the essential mechanisms involved in the determination of the behavior of the interacting populations. For this reason, it is felt that we can gain understanding by analyzing our models. For example, in the prey-predator interaction we were able to determine what happened to each population under the assumption that the prey and the predator populations affected each other in a definite and prescribed manner. Now we hoped that the prescribed effect of interaction approximated the actual interaction and hence by analyzing the results of our program runs, we can or cannot confirm the validity of our hypotheses. In this way, insight and understanding of the actual interaction of the populations is gained. Of course, if we are fortunate enough that, for a wide range of values of the parameters, our results very closely approximate actual observed behavior, we then have some faith in using the model as a predictive tool. Using models for prediction is fraught with pitfalls; nevertheless in making decisions of any sort involving future occurrences, some kind of model must be used. The principal reason that the results of computer based models are subject to such severe scrutiny is (the very fact) that the basis of these models can be laid bare for all to see. Thus, there is not room for rhetoric, vague phrases, and other methods of conviction that are so frequently used when presenting an argument. The computer program is very specific and hence the prescription for calculating the results is clearly evident in the program. For this reason, it is possible to discern what the program is doing and to consequently know the actual basis upon which the results were obtained. There is no ambiguity or vagueness in the programming statements and thus there are no hidden agendas.

This brief discussion concerning the philosophy of modeling was given to again remind the student that the importance of a model is the ability to test and analyze the consequences of the assumptions upon which the model is constructed. It is sometimes possible, as we have seen above, to arrive at the same model using different assumptions. Therefore, in the light of the previous discussion, it is important to interpret and analyze the results of the model always recognizing the particular assumptions from which the model was constructed.

The kinds of equations that we have been considering in these first two chapters are called finite difference equations. A very readable discussion of such equations is given in the text by Goldberg (1958). It is known that there exist finite difference equations which are unstable; that is, when they are programmed and run on a computer the numerical results bear no relation to the true results. This is because of the limited arithmetic precision of the computer. Speaking very loosely, such behavior is called numerical instability and is discussed in courses in numerical analysis.

PROBLEMS

CHAPTER II

1. Modify the two independent growth equations to include:
    (a) The effects of a finite food supply and prey-predator interaction.
    (b) The effects of contamination and a finite food supply.
    (c) The effects of contamination assuming an abundance of food, space, etc. and no prey-predator interaction.

    List, and clearly define, all notations.
2. Write a program to analyze 1(a). Choose values for the parameters and make some runs. Discuss your results.
3. In a two population finite resource model, the natural growth rates for the first and second populations are 0.4 and 0.6 respectively. Let the initial populations be 100 and 200 and let the constants R1, R2, T1 and T2 have the respective values 0.001, 0.002, 0.3 and 0.2.
    (a) Determine which population will survive.
    (b) How many generations before the dying population becomes less than 1, i.e. dies out.
    (c) Make some runs, using different initial populations and record the number of generations before one population dies out. Plot the results. Discuss them.
    (d) Make some runs with different values of G1 and T1. Graphically display the number of generations required for the vanishing of a specie versus the ratio G1/T1. Discuss the results.
4. An insect population is increasing 10 percent per week during the summer months of April 1 through October 1. In the remaining or winter months, the population decreases 5 percent per month. A migratory bird population resides in the same habitat as the insects during the months of May through August. Each bird eats approximately 50 insects per day and during the month of June the birds reproduce. It is assumed that corresponding to each parent, two newborn survive to become adults. It is further assumed that by the end of June the newborn birds are

eating as many insects as the adult birds and that during the month of June they required no insects as food. Finally, it is assumed that during the remaining months from September through April, the mortality rate of the birds is 60 percent. If the initial population of the insects is 100 million, and the initial population of the birds is 1000 on May 1, graphically display the time evolution of each population for a period of five years. Use a time increment of one day in your problem and assume there are 28 days each month

5. Two populations, P1 and P2, are coexistent. The daily change in the first population is directly proportional to the number of the second population while the daily change in the second population is such as to decrease the second population by an amount proportional to the second population. Describe the growth of both populations.

6. In a community of two species, the first species grows at the rate of 20% per month (30 days) whereas the second species grows at the rate of 5% per week. The daily food consumption of the first population is proportional to the first population with a constant of proportionality equal to 1 food unit per day per 1000 individuals. Similarly, the daily food consumption of the second population is proportional to the second population with a constant of proportionality equal to 2 food units per day per 1000 individuals. The effect of this food consumption on the growth rate of each population is such as to decrease the growth rates of each population by amounts proportional to the total food consumed each period. The respective constants of proportionality are 0.01 and 0.002. Describe the growth of the two populations. (HINT: One day should be chosen as the fundamental time period).

7. Using the conditions of problem 6, describe the growth of the two species assuming that the daily food consumption of each population is proportional to the square of each populationg with the same constants of proportionality.

8. In the derivation of the prey-predator program, it was assumed that the proportions of births and deaths of the prey were modified in proportion to the population of the predator. If instead, the change in the birth and death proportions of the

prey are assumed to be proportional to the cube of the predator population, describe the time evolution of both the prey and the predator.

9. Construct a program describing the time behavior of three populations assuming that two populations are prey for the third. State your hypotheses clearly. Describe the results.

10. Same as problem number 9, only assume that the first population is prey for the second and that both the first and second populations are prey for the third population

11. Construct a program describing the behavior of these populations competing for the same food supply. Describe your results.

REFERENCES

CHAPTER II

Emlen, J. M., 1973. Ecology: An Evolutionary Approach. Addison-Wesley, Reading, Mass.

Goldberg, S., 1958. Introduction to Difference Equations. J. Wiley & Sons, Inc., New York, NY

Innis, G. W., 1973. Dynamic Analysis in "Soft Science" Studies: In Defense of Difference Equations, pp. 103-121. Lecture Notes in Biomathematics. Ed. by S. Levin. Vol. 2. Springer Verlag.

Leslie, P. H., 1948. Some Further Notes on the Use of Matrices in Population Mathematics., pp. 213-245. Biometrika, Vol. 35, Parts III and IV. Dec.

Leslie, P. H., 1958. A Stochastic Model for Studying the Properties of Certain Biological Systems by Numerical Methods, pp. 16-31. Biometrika, Vol. 45, Parts I and II. June.

Leslie, P. H. and Gower, J. C., 1960. The Properties of a Stochastic Model for Two Competing Species, pp. 316-330. Biometrika, Vol. 45, Parts III and IV, Dec.

Muench, H., 1959. Catalytic Models in Epidemiology. Harvard Univ. Press.

Murdock, W. W. and Daten, R. V. 1974. Advances in Ecological Research. Vol IX, Academic Press, London and New York.

Murdock, W. W. and Daten, A. Population and Predation Stability. In Advances in Ecological Research, pp. 1-130, Vol. 9, 1975.

Poole, R. W., 1974. An Introduction to Quantitative Ecology. McGraw-Hill Co., New York, NY

CHAPTER III

PARAMETER DETERMINATION

Introduction

In the Malthus model the growth coefficient G was an undetermined parameter, whereas in the finite resource model both G and G1 were undetermined or free parameters. The student will recall that these parameters resulted from the assumption that one variable was proportional to another variable and that the parameters either were the constants of proportionality or were directly related to the constants of proportionality. In fact, G was defined as the difference between the birth rate B, and the mortality rate M, and G1 was defined in a similar manner. The determination of such parameters is frequently effected by a comparison of model results with experimentally obtained results. Sometimes it is possible to conduct experiments which will permit the determination of each parameter in a sequential and direct manner. It is also occasionally the case that some of the parameters may be obtained from calculations utilizing both theoretical and experimental results. Usually, however, the determination of the parameters is more difficult and it is not possible to determine each parameter sequentially. Consequently, a set of experimental data must be obtained and a comparison made with a set of computer or theoretical results in order to obtain the model parameters. The process is both lengthy and difficult and has necessitated the development of elegant mathematical and experimental techniques.

There are two primary sources of error which make the determination of the parameters difficult. They are (1) errors in the model, and (2) errors in the experimental data. The model can be in error because it was constructed on the basis of an incomplete or an incorrect hypothesis, because computational resource limitations prevented the construction of a more complete model, or because of programming errors. These latter errors are very difficult to detect since there rarely exists any known answers or data with which the program can be checked. The experimental data may be in error because of limitations in experimental technique, the impossibility of performing the exact desired experiment, inaccuracies in the recording of the data or just plain experimental blunders. All of these factors must be considered when the investigator is determining the model parameters. It is not our place to consider these sources of difficulty here except to say that in accordance with Murphy's Law, programming errors and experimental blunders usually occur and necessitate an attitude of eternal vigilance to prevent their occurrence. We will "assume away" all of the others and carry on as if they could not happen to nice people like us.

An Example

As the first example of the general problem of parameter determination, we consider the problem of determining the growth coefficient in the Malthus model. This problem is chosen because of the simplicity of presenting the ideas. In actual fact, it is quite difficult to experimentally insure such a population growth because of the difficulty of ensuring the hypotheses of a completely constant environment over a lengthy period of time.

The output of the computer program consists of a sequence of points which describe the time evolution of the population. If a smooth curve is drawn through these points, the resultant curve describes the population at any instant of time. This rather obvious statement is made because in the following we will use the terms, graph, curve, and sets of points interchangeably. As the student has noted, the graph changes shape as the growth parameter G changes. This suggests the possibility of determining the growth parameter by successive alterations of an initial estimate of its value until a population curve is obtained which coincides with, or is close to, the experimentally determined population curve. There are several alteration processes which have been used with varying degrees of effectiveness. The first process to be described is a trial and error process and is admittedly not elegant. Nevertheless, such a procedure is usually quite effective, especially when the investigator has some previous knowledge concerning some or all of the parameters. Moreover, since it depends upon the interaction of the investigator with the computer, the technique enables the investigator to develop some insight into the problem.

In multiple parameter determination problems, as well as in the single Malthus parameter determination problem, if a mathematical formulation had been employed to describe the phenomena, the solutions of the resulting equations would have contained the model parameters and the problem of the determination of these parameters would still have remained. This problem has a long mathematical history and many techniques have been developed to facilitate its solution. These techniques are usually iterative procedures, i.e. they require an initial guess at the value of the parameters and then a "correction process" is applied which hopefully, will improve the initial guess. The corrected set of values is used as the initial guess for the determination of the next set of values and the process repeated. The procedure is continued until the theoretically obtained graph is suitably close to the experimentally obtained curve and then the parameter values so required are assumed to be the desired values. There is thus, a very close similarity between the model approach and the mathematical approach to the problem.

Perfect agreement, that is complete coincidence of the theoretical and experimental data, cannot be expected and hence, we will have to be satisfied with those parameter values yielding a calculated curve which is close to the experimental curve. Thus, we are making the tacit assumption that large changes in the values of the parameters will produce correspondingly large changes in the resulting population curve because this will insure that two very different sets of parameter values would not produce two population curves equally close to the experimental curve. If two widely different sets of parameters did produce almost coincident population curves, it would be most difficult to accurately determine the parameters. We would say that the model is very insensitive to these parameters and that quite possibly the model could be improved or some other technique should be used to obtain the parameters.

Comparison of Two Curves

The criteria for selecting the model parameters depends upon the ability to determine the comparative closeness of different computer generated curves with the experimentally generated curves. In particular, if we have two curves labeled C_1 and C_2, how can we decide which curve is closer to a third curve labeled, C_3. If the three curves were three points and the points labeled P_1, P_2, and P_3, the problem would have a simple solution. We would use a ruler and just measure the distance between the two points P_1 and P_3 and the distance between the two points P_2 and P_3, and then compare the results. However, the obtaining of a numerical estimate for the distance between two curves is a more difficult task since we do not have a correspondingly simple method of numerically estimating the distance between two curves. We are thus faced with the problem of trying to establish some way of assigning a numerical value to the distance between two curves. In mathematics, such an expression is called a Norm.

The establishment of a suitable measure for the numerical estimate of the distance between two curves is greatly facilitated by examining a pictorial representation of the problem. Figure 3.1 below indicates a portion of two curves labeled C_1 and C_2 respectively.

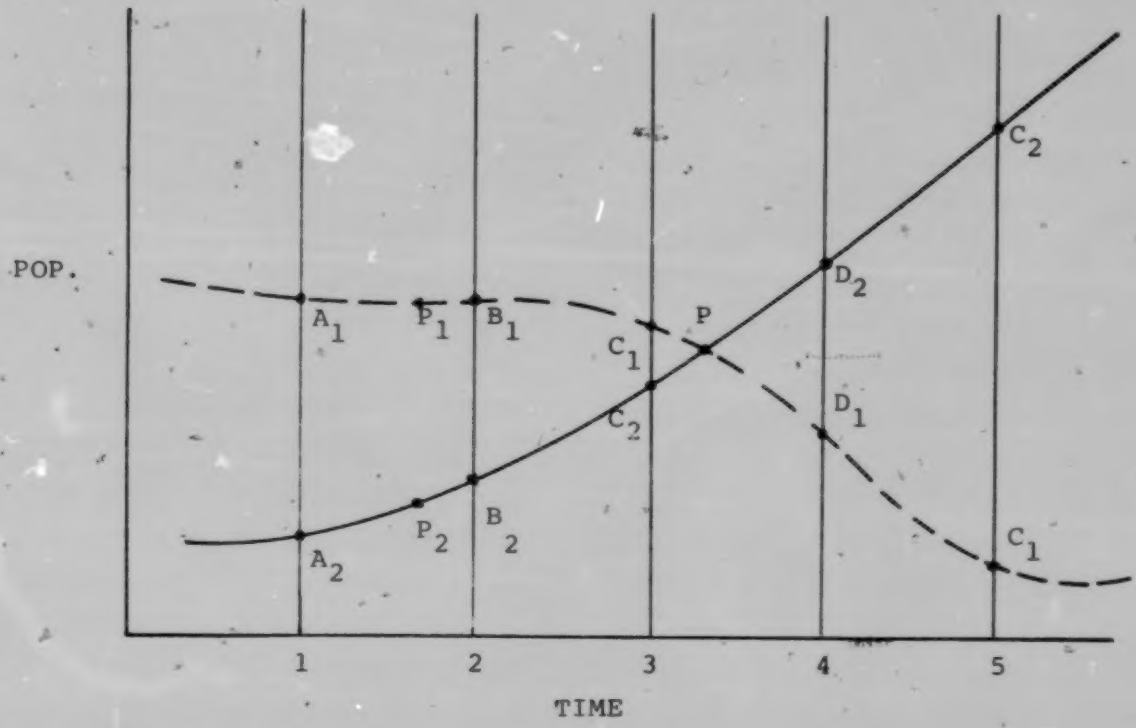


Figure 3.1

The points A_1, B_1, C_1, and D_1, belong to the curve C_1, and the points A_2, B_2, C_2, and D_2, belong to the curve C_2. Before attempting to formulate an expression for the numerical estimate of the distance between the two curves, it is necessary to establish some desirable characteristics that such an expression should possess. An examination of the figure will suggest some of these characteristics. First of all, the estimate should have the property that a large numerical value of the distance estimate should correspond to the fact that the curves are far apart, and conversely, a small numerical estimate should imply that the curves are close together. In particular, if the magnitude of the numerical estimate for the distance is zero, the two curves should coincide, and conversely; if the two curves coincide, the numerical estimate should be zero. A second desirable feature for the expression for the estimate of the distance is that the estimate for the distance between the curves C_2 and C_1 should be equal to the distance estimate between the curves C_1 and C_2. Finally, the numerical estimate of the distance should also possess the property that the numerical estimate of the distance between the two curves C_1 and C_2, when added to the distance estimate between the two curves C_2 and C_3, should be greater than or equal to the numerical estimate of the distance between curves C_1 and C_3. This is a generalization of the fact that a straight line is the shortest distance between two points. With these criteria in mind we proceed to "guess at" or "cobble up" an expression for the numerical determination of the distance between two curves. The fact that we can numerically determine the distance between two points and the idea that we can imagine each curve as made up of a large number of points suggests a method of determining the required estimate.

In order to make the subsequent discussion more easy to follow, it is convenient to introduce the following definition. Let P_1 be any point on curve C_1, and P_2 be any point on curve C_2, then the pair of points (P_1, P_2) are said to be corresponding points if they have the same abscissa or horizontal coordinate. Thus, if a pair of corresponding points have the same ordinate or vertical coordinate, the points coincide and this point is a point of intersection of the two curves. Furthermore, if a section of the first curve is close to a section of the second curve the differences of the ordinates (these differences are called deviations) of corresponding points on these sections is small. These facts suggest that a summation of the differences of ordinate values for several corresponding points should provide a reasonable measure of the distance between the two curves. This estimate is defective however, since an examination of the corresponding pairs of points (B_1, B_2) and (D_1, D_2) in Fig. 3.1 reveals that the difference in vertical values of B_1 and B_2 is the negative of the difference of the vertical values of D_1 and D_2. Hence, the sum of these deviations is zero and yet the portions of the curves subtended by these points are certainly not coincident. This defect is readily corrected by always taking the difference between the ordinate values as positive, i.e. taking the absolute value or magnitude of the difference. It is then seen that if the sum of the absolute value of the deviations of a specified set of corresponding pairs of points is zero that, at least at these points, the curves must coincide. Furthermore, the distance as calculated gives the same value for the distance from C_1 to C_2 as from C_2 to C_1. Finally, if the sum is very small for a large number of points and the curves are "smooth" and do not have many wiggles, then the curves will be close together. If the BASIC programming language notation is used to indicate the absolute value of a quantity, the sum of the absolute values of the difference of the ordinate values would appear as

$$\mathrm{ABS}(A_1-A_2) + \mathrm{ABS}(B_1-B_2) + \mathrm{ABS}(C_1-C_2) + \mathrm{ABS}(D_1-D_2).$$

In this expression $A_1, A_2, B_1, \ldots, D_2$ indicates the vertical values or ordinates of the points.

A very popular method of numerically estimating or determining the distance between two curves is to calculate the sum of the squares of the deviations. With reference to Fig. 3.1 and the four points shown, this criteria may be expressed as

$$(A_1-A_2)^2 + (B_1-B_2)^2 + (C_1-C_2)^2 + (D_1-D_2)^2$$

where the subscripted variables again indicate the ordinate or vertical values of the points.

Another very popular distance criteria is the maximum magnitude of the deviations. Using the same illustrative example, this criteria is obtained by first calculating the four quantities

$$\mathrm{ABS}(A_1-A_2), \qquad \mathrm{ABS}(B_1-B_2), \qquad \mathrm{ABS}(C_1-C_2), \qquad \text{and} \qquad \mathrm{ABS}(D_1-D_2).$$

and then determining the largest of these. This estimate is sometimes called the Max. Norm. and measures the closeness of two curves by the magnitude of the largest distance between pairs of corresponding points. It is seen that all the properties of a distance measure are fulfilled and in particular, if the maximum deviation is zero, the curves will coincide.

Still another method is based upon a calculation of the area between two curves, being careful to consider all such areas as positive. Each of these methods satisfies all of our desired criteria for the numerical estimation of the distance between two curves. In this work, the criteria for closeness will frequently be chosen to be the sum of the squares of the deviations since this criteria is the easiest to calculate. The student may be familiar with such a criteria from courses in statistics where the method of "least squares" plays such an important role. Thus, in the work below we will calculate the sum of the squares of the differences of several pairs of corresponding points on the two curves and if this difference is small, we will conclude that the curves are close together. If a pair of curves is a finite, though small, distance apart we note that by increasing the number of pairs of corresponding points we can indefinitely increase the value of the sum and this would imply that no matter how close the two curves were that by taking enough pairs of points in the calculation we could always conclude that the curves are far apart. However, the criteria for the determination of the distance between two points is to be used to compare the degrees of closeness of two curves, C_1 and C_2, to curve C_3 and thus we will choose the same number of pairs of corresponding points in each comparison. This criteria also requires that the points chosen on the C_1 and C_2 curves correspond to the points chosen on the C_3 curve. In this manner the smaller sum will provide a consistent measure for determining which of the curves, C_1 or C_2, is closest to curve C_3.

Comparison with Experiment

Having established a usable measure that will permit the comparison of both model and experimental results, we return to the problem of determining the growth parameter G in the Malthus model. The recent growth of many human populations has been described as exponential. In particular the United States, with its vast area and relative abundance of natural resources, has provided an approximate constant environment for the growth of the population up until recent times. Thus, the recorded population growth of the United States should provide somewhat realistic data to use as a basis of comparison. The changes in population due to immigration will be ignored and it will be assumed that the Malthus model results can approximate the actual population growth. The growth is presented in Table 3.2 and the data has been taken from Peterson (1961).

The experimental data is presented for equal increments of time; however, it is often the case that such data is obtainable only at unequal time increments or at instants of time which do not coincide with the computer results. Since the criteria for closeness requires that pairs of experimental and theoretical points be corresponding points, interpolation or curve fitting techniques must be used to obtain data that enables the use of corresponding experimental and theoretical points. Such techniques will be discussed later (see Chapter IV), and thus the present example will use the same time period in the model as employed in the experiment. This will insure the existence of corresponding data points.

A Suggested Method

We will determine G using the sum of the squares of the deviations as the criteria for the closeness of two curves. Hence, we will modify the constant resource model computer program to calculate the sum of the squares of the differences between pairs of corresponding experimental and calculated points. With the aid of this program, the

TABLE 3.2

U. S. Population Growth

Year	Population (Millions)
1790	3.93
1800	5.31
10	7.24
20	9.64
30	12.87
40	17.07
50	23.19
60	31.44
70	39.82
80	50.16
90	62.95
1900	75.99
10	91.97
20	105.71
30	122.78
40	131.67
50	150.70
60	179.32

determination of G will be accomplished by finding that value of G which results in the minimum value for the sum of the squares of the deviations. The time increment in the model will be assumed to be ten years to coincide with the time increment in the experiment. This implies that the growth coefficient, when so obtained, is with respect to a ten-year period. Thus, we would be obtaining the relative change in the population over a ten year period instead of over the normal single year period. It is possible to calculate the populations assuming a single year period and to then compare the populations every ten years. See problem number 1 at the end of the chapter. The results are somewhat different than what might be expected. The growth coefficient so obtained is a yearly growth coefficient, that is, the relative change in the population each year.

The procedure for determining G will be very heuristic. It will consist of assuming different values for G and then using the program to evaluate the sum of the squares of the deviations, S, of the pairs of corresponding computational and experimental points. The value of G resulting in the smallest sum will be assumed to be the true value. This is a pure trial and error procedure. Hopefully, however, our intuition will be increased as the number of trials increases and thus the process should not be too lengthy. Our intuition is usually always greatly increased when such a procedure is used and this is one of the principle benefits of such a very heuristic procedure. Complicated or elegant methods sometimes render the obtaining of answers easier, but render the obtaining of insight more difficult. In actual practice, usually only a small number of trials are necessary to obtain fair results. This last statement may require some clarification. In essence, the problem is to know when to stop correcting the value of G. One criteria for stopping is exhaustion, but this is hardly a workable criteria. The process is usually stopped when it appears that small changes in G produce only very, very small changes in the sum. In order that the trial and error process be effective, it is assumed that the model approximates the experimental phenomena sufficiently closely so that there truly does exist only one minimum point, i.e only one value of G for which the sum of the squares of the deviation is a minimum. The perceptive or mathematically sophisticated student will note that we are assuming there are no relative minimum points for the surface S.

A principal value of the guessing, or trial and error technique, is that complete failure of the method frequently serves to indicate a significant deficiency in the model. Furthermore, for very complicated models such a technique may very well be the only feasible technique. Finally, it may be the case that there are some exterior imposed constraints on the parameters, e.g. each should be larger than some value, yet the sum is not a minimum for parameter values satisfying these constraints. Thus, the investigator must compromise, which compromise must partially be determined by conditions exterior to the problem as well as the desire to minimize S. In almost all cases, a satisfactory compromise is obtainable only with the assistance of the intuition and objectives of the investigator. It is only fair to point out to the student that the assumption that there does indeed exist a minimum point, and that furthermore, this point is the only such point, is a very strong assumption. In practice, this frequently is not the case, and there do exist other relative minimum points. Since the purpose of this section is to illustrate a method which is often times successful, a discussion of alternative methods, improvements and pitfalls will be postponed until later.

The Program Modifications

The original Malthus model program has been modified to calculate the sum of the squares of the deviations. Because it was very easy to do, the program was also modified to determine the deviation of largest magnitude as well as the magnitude of this deviation. In addition, the program was altered to also calculate and print the relative error corresponding to each observation point. The relative error, R, for a period is defined to be the deviation, D, between the theoretical and the observed population at this period divided by the observed population. Thus, since $D = P(D)-E(I)$, it follows that $R = D/E(I)$. As mentioned previously, the numerical value of the deviation of largest magnitude and the distribution of the magnitudes of the relative errors are other estimates of the degree of closeness of two curves. They are included in this program because their calculation requires very little extra effort. The program is listed in Fig. 3.3 and

```
1 REM      SINGLE VARIABLE CURVE FITTING
2 REM      SIMPLE POPULATION GROWTH MODEL
3 REM   LEAST SQUARES AND MIN MAX FOR DEVIATIONS
5 DIM P(50),E(50)
10 GOSUB 205
15 PRINT "INPUT  G,  P(0)"
20 INPUT G,P(0)
25 REM   INSTR. NOS. 30 TO 40 CALCULATE THE POPULATIONS
30 FOR I=0 TO 25
35 LET P(I+1)=P(I)+G*P(I)
40 NEXT I
41 PRINT
45 REM INST. NOS. 50 TO 85 CALCULATE THE MEASURES OF CLOSENESS
46 PRINT
50 LET S=0:LET K=0
55 LET M=ABS(P(0)-E(0))
57 LET R=M/E(0)
60 PRINT K,P(0),E(0),M,R
65 FOR I=1 TO 17
70 LET D=ABS(P(I)-E(I))
75 IF D<=MGO TO 85
80 LET M=D:LET K=I
85 LET S=S+D*D
90 LET R=D/E(I)
95 PRINT I,P(I),E(I),D,R
100 NEXT I
105 PRINT
110 PRINT "THE VALUES OF  G,  S,  M  AND  K  ARE"
115 PRINT G,S,M,K
120 STOP
155 RETURN
160 END
200 REM    INSTR. NOS. 205 TO 230 ENTER THE EXPERIMENTAL DATA
205 DATA 3. 93,5. 31,7. 24,9. 64,12. 87,17. 07,23. 19,31. 44
210 DATA 39. 82,50. 16,62. 95,75. 99,91. 97,105. 71
215 DATA 122. 73,131. 67,150. 7,179. 32
220 FOR J=0 TO 17
225 READ E(J)
230 NEXT J
235 RETURN
240 END

READY
```

Figure 3.3

the three measures of closeness are calculated in instructions 50 to 100. Instructions 30 to 40 calculate the population using the assumed value of the growth coefficient and the experimental data, E(I), is entered by instructions 205 to 230.

The program can be used to assess, by different criteria, the closeness of the theoretical and experimental data. The use of more than one criteria for the closeness of two curves is presented to illustrate the fact that different parameter values may be obtained depending upon which closeness criteria is used. As previously stated, there are many criteria for the closeness of two curves. In fact, if the relative error is small in magnitude for all observation points of interest, this indicates that the curves are close together over the entire time period or range. In contrast, the fact that the sum of the squares of the deviations is small indicates that the total variation between the two curves is also small. It is possible for the total variation to be small and yet for the relative error at a single or a few points to be quite large. Consequently, the uniform smallness of the relative error together with a comparatively small, or minimum value, for the sum of squares of the deviations provides a quite good working criteria for testing the closeness of the two curves.

Your author has used this program to determine the "best" value for G and detailed results are presented in Tables 3.4, 3.5, and 3.6. Table 3.4 gives a detail of a "typical" run. The column headed I indicates the decade number and the columns labeled P(I) and E(I) list the calculated population and the observed population respectively for each decade. The last two columns, entitled D and E list the deviation, (P(I)-E(I)) and the relative error (P(I)-E(I))/E(I) respectively. At the bottom of the table there is given the value of the growth coefficient G used to calculate the population, the value of the sum of the squares of the deviations, S, the magnitude, M, of the largest deviation, and the decade in which this deviation occurred. The values of A=0.265 and P(0)=3.93 million were used in this run and a graphical comparison of the calculated population growth versus the known or experimental growths is given in figure 3.7. In Tables 3.5 and 3.6, the results appearing in the columns labeled S and M are rounded figures and the number listed in the column headed K is the decade period corresponding to the deviation of largest magnitude. In each of these tables, the results of a dozen or so runs are

Table 3.4

Results of a Typical Run
Using Program in Fig. 3.3

```
RUN

RB002

INPUT  G,  P(0)
?.265,  3.93


I              P(I)         E(I)          D            R

0              3.93         3.93          0            0
1              4.97145      5.31          .33855       .0637571
2              6.28888      7.24          .951116      .13137
3              7.95544      9.64          1.68456      .174747
4              10.0636      12.87         2.80637      .213055
5              12.7305      17.07         4.33951      .254218
6              16.1041      23.19         7.08593      .30556
7              20.3717      31.44         11.0683      .352047
8              25.7701      39.82         14.0499      .352834
9              32.5992      50.16         17.5608      .350095
10             41.238       62.95         21.712       .344908
11             52.1661      75.99         23.8239      .313514
12             65.9901      91.97         25.9799      .282482
13             83.4775      105.71        22.2325      .210316
14             105.599      122.73        17.131       .139583
15             133.583      131.67        1.91278      .0145271
16             168.982      150.7         18.2822      .121315
17             213.763      179.32        34.4425      .192073

THE VALUES OF  G,   S,   M   AND  K   ARE
 265           4734.94      34.4425       17
READY
```

Run	G	S (Sum of Squares of Devs.)	M (Max Mag of Dev.)	K (Time Period for M)
1	.265	4734.9	34.44	17
2	.27	5838.1	49.27	17
3	.26	4597.3	29.04	12
4	.25	6622.9	34.78	12
5	.2	37770.94	92.13	17
6	.3	44429.4	160.64	17
7	.263	4573.1	28.77	17
8	.262	4546.807	27.83	12
9	.261	4555.28	28.44	12
10	.2615	4546.775	28.14	12
11	.2613	4549.16	28.26	12
12	.2618	4545.76	27.96	12

Table 3.5

Comparison of Results of Several Runs
(Initial Population Equal to 3.93 Million)

Run	P(0) (Init. Pop.)	G	S (Sum of Squares of Devs.)	M (Max Mag of Dev.)	K (Time Period for M)
1	3.5	0.262	6,462.1	34.85	12
2	3.8	0.262	4,839.9	29.95	12
3	4.5	0.262	6,186.77	55.76	17
4	4.2	0.262	4,729.8	40.09	17
5	4.1	0.262	4,537.4	34.87	17
6	4.0	0.262	4,491.6	29.64	17
7	4.1	0.261	4,423.65	32.00	17
8	4.2	0.261	4,539.4	37.15	17
9	3.8	0.263	4,776.37	29.36	12
10	3.9	0.263	4,597.5	27.7	12
11	3.7	0.263	5,105.5	31.01	12
12	3.7	0.264	5,007.1	30.43	12
13	3.8	0.264	4,745.46	28.77	12
14	3.9	0.264	4,637.70	29.97	17

Table 3.6

Comparison of Results of Several Runs
(Different Initial Populations)

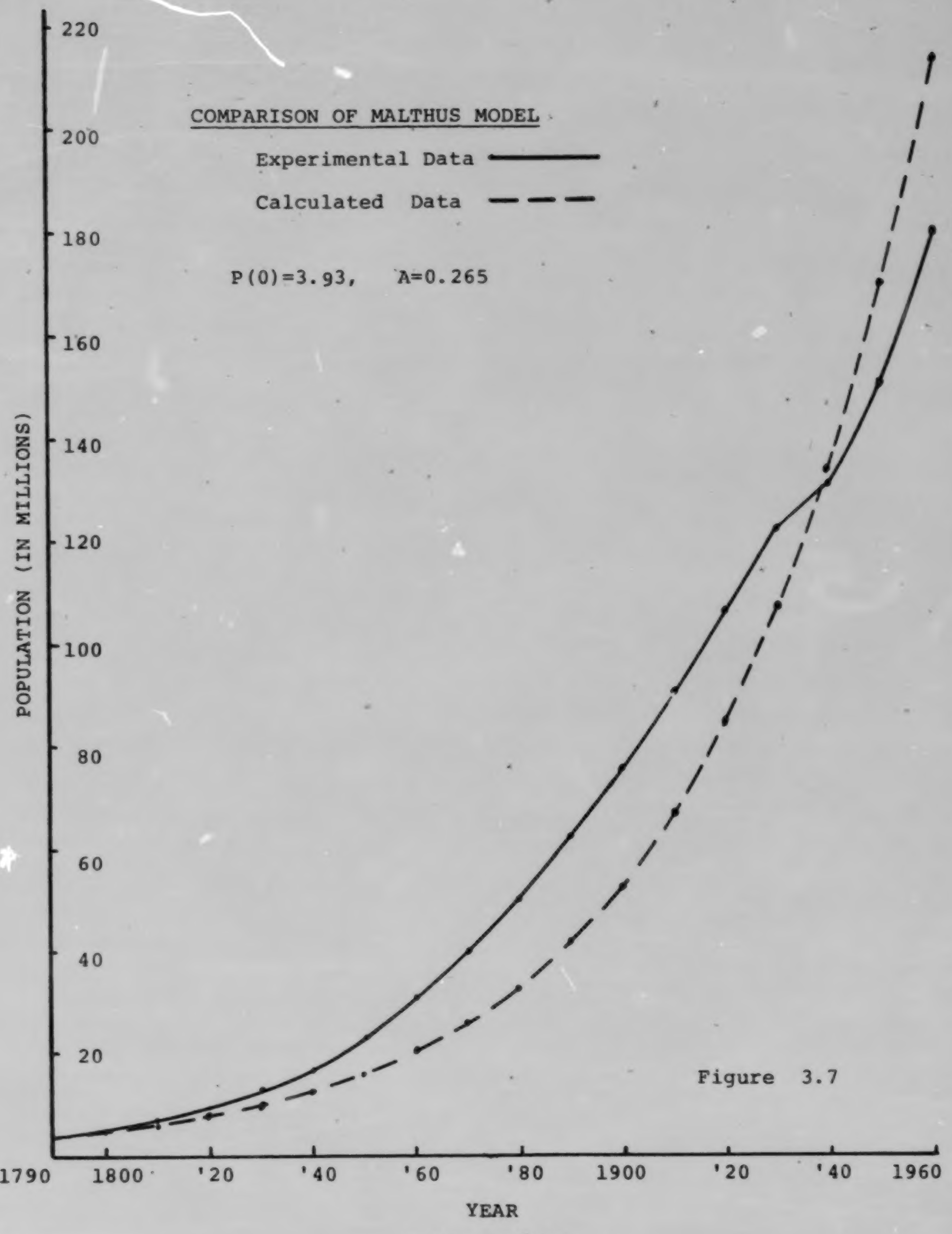

COMPARISON OF MALTHUS MODEL
Experimental Data
Calculated Data
P(0)=3.93, A=0.265
POPULATION (IN MILLIONS)
220
200
180
160
140
120
100
80
60
40
20
1790
1800
'20
'40
'60
'80
1900
'20
'40
1960
YEAR

Figure 3.7

presented. Table 3.5 presents results obtained by using an initial population identical to the experimental population of 3.93 million persons. Table 3.6 presents results wherein the initial population is varied slightly around the initial experimental value. By varying the initial population, a different value for the growth parameter as well as a smaller value for the sum of the squares of the deviations is obtained. Hence, we conclude that if the sum of the squares of the deviations is used for the closeness criteria that the best results are obtained by varying both the growth coefficient and the initial population. Because of exterior constrants, it may be the case that the initial population must coincide with the observed value. In this event, on the basis of the results shown in Table 3.5, the best value for G is 0.2615. On the other hand, if the investigator decides that it is permissable to vary the initial population, the results of Table 3.6 imply that with an initial population of 4.1 million, the best value for the growth coefficient is 0.261. The student should understand that the decision of whether or not to permit the variation of the initial population is not a mathematical decision. The decision must be made by the investigator and be in accord with his objectives. Further examination of the results in Tables 3.5 and 3.6 also reveals that small changes in the parameter values did sometimes produce significant changes in the value for the sum, S, as well as for the numerical value of the deviation of largest magnitude. This fact implies that the aforementioned criteria for discontinuing the iteration must be applied with some care.

Table 3.5 also illustrates that, for the runs shown, the minimum value of the maximum deviation is obtained for G = 0.262, whereas the smallest sum of squares of the deviations is obtained for G = 0.2615 Table 3.6 reveals a similar behavior. The value of G equal to 0.263 corresponding to an initial population of 3.9 million results in the minimum-maximum deviation. This value is slightly different than the least squares value of 0.261 corresponding to an initial population of 4.1 million. It is usually the case that different parameter values are obtained for different criteria of closeness.

The resolution of the question of which measure of closeness to use is a problem to be decided by the investigator and is not a mathematical one. Recently, the idea of selecting the parameter value which results in the minimum value for the maximum deviation has become very popular in mathematical approximation theory and is being used quite extensively in this field. This is called the min max criteria. As stated previously, there also exist other closeness criteria. It is well for the investigator to understand the limitations of each measure, because he must substantiate his choice, interpret his results, and infer his conclusions accordingly. We will discuss all of these problems again in the next chapter.

In the previous example, the value of G was obtained by using a hit and miss process to find a parameter value which would result in a minimum value for the particular closeness criteria. It is important to note that the value of G so obtained is not necessarily the value that would produce the absolutely lowest sum. To find such a value, a prescribed iterative process would have to be constructed and the process proved to yield the parameter value corresponding to the minimum value of the chosen closeness criteria. This analysis will not be carried out here and is best left to another more advanced course. The value to the investigator of such a proof should not be underestimated since it would assure him that he had indeed achieved the value of G resulting in a minimum for the chosen closeness criteria. In the example, the sum of the squares of the deviations was chosen as the criteria for closeness. The trial and error process was stopped after obtaining what seemed to be two figure accuracy recognizing that the very restrictive hypotheses in our model probably made the continued refinement of the parameter value unrealistic. It is always well to keep in mind that the program results are obtained from hypotheses which do not completely and accurately describe the phenomena under investigation. Consequently, there is a realistic limit to the accuracy of such a program and to use the program to determine the parameters to a large number of significant figures is nonsense. In addition, the observed data contains experimental errors and hence, excessive accuracy is again unwarranted.

For this simple example, it is also feasible to plot the two curves and visually compare the results. Such a procedure does not appear to be very scientific; nevertheless, it is a very useful one. Visual observation and comparison is of great assistance in obtaining insight. Complex programs with several tables of results usually prevent the effective plotting of all results. Even a moderately complex computer program can generate a vast amount of output and the analysis and effective use of this information can be a staggering task. It is also true that far more useful information than can be readily outputted is generated by such programs. Thus, it is frequently the case that decisions concerning which results to display, as well as the development of methods of analysis of the results, are at least as difficult a task as the development of the original program.

The previous work is very heuristic and has been presented in such a manner as to provide the student with an overall intuitive feel for the problem. A rigorous analysis of techniques for determining the closeness of two curves is the subject of approximation theory and is beyond the scope of this work. We reiterate, the purpose of this work is to encourage the student to think quantitatively and to develop a sense of the power and ready applicability of the digital computer. Consequently, in this introductory course, heuristic methods and a cavalier (albeit, hopefully honest) attitude is encouraged. The serious student will then recognize that a more careful analysis of the procedures is needed and he is referred to texts in numerical analysis, approximation theory and statistics. We will touch on some of these matters later, but only very briefly.

PROBLEMS

CHAPTER III

1. Using the population data for the United States as given in Table 1, obtain the yearly growth coefficient, G, for the Malthus model and the least squares norm by (a) starting with the known population, and (b) by varying the initial population.
2. Using the same data and model as in problem #1, obtain the growth coefficient for the min max norm.
3. Same as problem #2 above, only use the min max norm applied to the relative error.
4. With the aid of a plotting or graphing routine, or by observing the appropriate calculated information, determine G so that the magnitude of the relative error is as nearly constant as possible.
5. Construct other measures of closeness. Write computer programs using each and compare and discuss the results.
6. Find the minimum value and the minimum point for

   (a) $A^2 + 10A + 15$

   (b) $A^4 + 4A^2 + 10$

7. Using the Malthus model and the tabular data listed below, determine the growth coefficient G using

   (a) The least squares norm.

   (b) The min max norm.

I	1	2	4	6	8	10	20
P	10	11.664	13.6049	15.8687	18.5093	21.58925	46.6096

CHAPTER IV

AUTOMATED PARAMETER DETERMINATION

Introduction

The previous chapter discussed and outlined a heuristic method for the determination of the parameter occurring in the constant environment model. Since such constants occur in every model and it is necessary to know such constants in order to use the model, their determination is of paramount interest to the investigator. Frequently, because of the inability to otherwise determine the parameters, they must be determined with the aid of the model itself, and the complexity of the model may render the heuristic method the only available method. The previously described minimization technique can also be applied, with very evident modifications, to the determination of parameters when the quantity is to be maximized rather than minimized. Finally, many important and diverse problems can be recast as minimization or maximization problems, and hence the technique has wide applicability. For these reasons, we extend our comments on the method and discuss some of the questions raised in the previous chapter. The comments will be limited to methods for the determination of a single parameter. In actual practice there are very few single parameter models. Even the constant environment model should be considered as having two parameters because, as we noted, the initial population can assume the role of a free parameter. Nevertheless, further discussion of the single parameter problem is warranted because it will provide much insight for the problem of the determination of more than one parameter. Two, or more, parameter problems will be discussed in the next chapter.

Minimization

The previous method for the determination of the parameter was based upon the premise that a unique value of the parameter would result in a specific quantity being the minimum value for this value of the parameter. In our example, the quantity was the sum of the squares of the deviations, and the previous statement means that any other value of the parameter would result in a larger value for the sum of the squares of the deviations. In order to avoid excess verbage it is convenient to introduce some notation. Let the specific quantity to be minimized be called the *criteria function* and denote it by M. The model parameter will be denoted by A. The value of A that results in the minimum value for the criteria function will be called the *extremum* or *extreme value*. Finally, the problem of determining the extreme value corresponding to a specified criteria function will be called the *minimization problem*. In the previous chapter, M was the sum of the squares of the deviations, S. M could also have been the maximum value of the relative error or the sum of the absolute magnitudes of the deviations, etc. The parameter A may be identified with the parameter G in our constant environment model. Using this terminology our problem can be restated as one of determining the value of A which makes M a minimum and the only requirement imposed upon M is that it be calculable. It should also be noted that in the example, the deviations were calculated at equal increments of time, i.e. ten years. The equality of time increments is not necessary and unequal time increments are frequently used because of the inability to obtain experimental data at uniform increments of time. This statement does not mean that the model results are to be obtained using unequal time increments; rather, it means that the calculation of M may use unequal increments of time. Since equal time increments were used to obtain the model results, the obtaining of model results at unequal time increments will necessitate the use of interpolation from the equal time increment data. This is discussed in a later section entitled, "Interpolation".

The minimization problem has a long mathematical history and its study has resulted in the creation of much beautiful mathematics. With the advent of computing machines there has appeared renewed interest on the part of mathematicians to obtain effective numerical algorithms for its solution. There is, however, a fundamental difference between the problem as we have considered it and the problem as the mathematician has traditionally viewed it. In the typical mathematically posed version of the problem, a family of functions and a set of parameters is given together with an expression involving both the functions and the parameters. The problem is to then determine the values of the parameters so that when the resulting functions are used to evaluate the expression, the expression assumes a minimum value. As an example, consider the problem of approximating the monomial X^4 by the expression $aX + bX^2 + cX^3$ for a known set of values of X. In this example, the family of functions is X, X^2, and X^3; the parameters are a, b, and c, and the expression M, to be minimized is

$$M = \sum_{i=1}^{10} [X_i^4 - (aX_i + bX_i^2 + cX_i^3)]^2 .$$

The symbol $\sum_{i=1}^{10}$ indicates that the square of the quantity in brackets is to be evaluated for the prescribed values X_1, X_2, X_3, etc. up to and including X_{10}, and the results totaled. The student will note that for each triple of values, (a,b,c), there exists one value for M. The problem then, is to determine the value of each of the parameters, a,b, and c, which render M a minimum. As this example shows, in the mathematically posed version of the problem, the family of functions was prescribed or given beforehand, whereas in our problem they are not given. In fact, they are not even known, because all that is given is a set of program statements whose execution by the computer result in a set of numbers which can be used to evaluate the criteria function which is to be minimized. This fundamental difference in the two problems means that much of the previous mathematical work is not readily applicable to our problem.

For those students who have had advanced mathematics, the problem as posed using the programming language BASIC is more closely related to the mathematical problem of determining by least squares, or some minimization technique, the parameters occurring in a differential equation when it is not possible to write down an explicit form of the solution of the differential equation and resort must be made to numerical methods of solution. This problem is recent and difficult and has also come into prominence with the advent of the digital computer. For these reasons there is a shortage of useful results and thus we cannot expect conventional mathematical analysis to be of great assistance in the solution of our problem. Nevertheless, the problem as posed in the BASIC language has a distinct advantage over the classical mathematical method of analyzing problems. In only rare instances is the investigator fortunate enough to be able to mathematically formulate the problem in such a manner that he can "capture the essence of the phenomena" and also solve the resultant model equation, or equations, in closed form. Because of this difficulty of obtaining a complete mathematical description of the phenomena, the investigator frequently avoids mathematical modeling altogether and resorts to straight curve fitting with polynomials. He then may attempt to attach some kind of meaning to the coefficients of the polynomial. Such an attempt is difficult at best because there is nothing biological or natural about the family of monomials 1, X, X^2, X^3, etc. They possess no inherent biological or natural characteristics. In contrast, an expression of the problem in a programming language results in a computer program which is a direct realization of the investigator's concept of the behavior of the phenomena and consequently the parameters enter the problem in a more natural way. Because of this, it is usually possible to associate a biological characteristic with them and hence, the task of gaining insight is much easier using the programming language formulation than using the mathematical formulation.

It is known that for conventional mathematically posed minimization problems that there may exist two or more relative minimum points. In terms of our notation this means that there can exist two or more distinct values of A (called critical or extremum points) for which M is a local minimum. Thus, there can exist two values of A, say A1 and A2, such that the value of M at A1 is less than the value of M at local or nearby points of A1 and also the value of M at the point A2 is less than the value of M at points in the near neighborhood of A2. The value of M at A1 need not be the same as the value of M at A2. The points A1 and A2 are called relative minimum points or local minimum points. Figure 4.1 indicates a minimization curve having three local

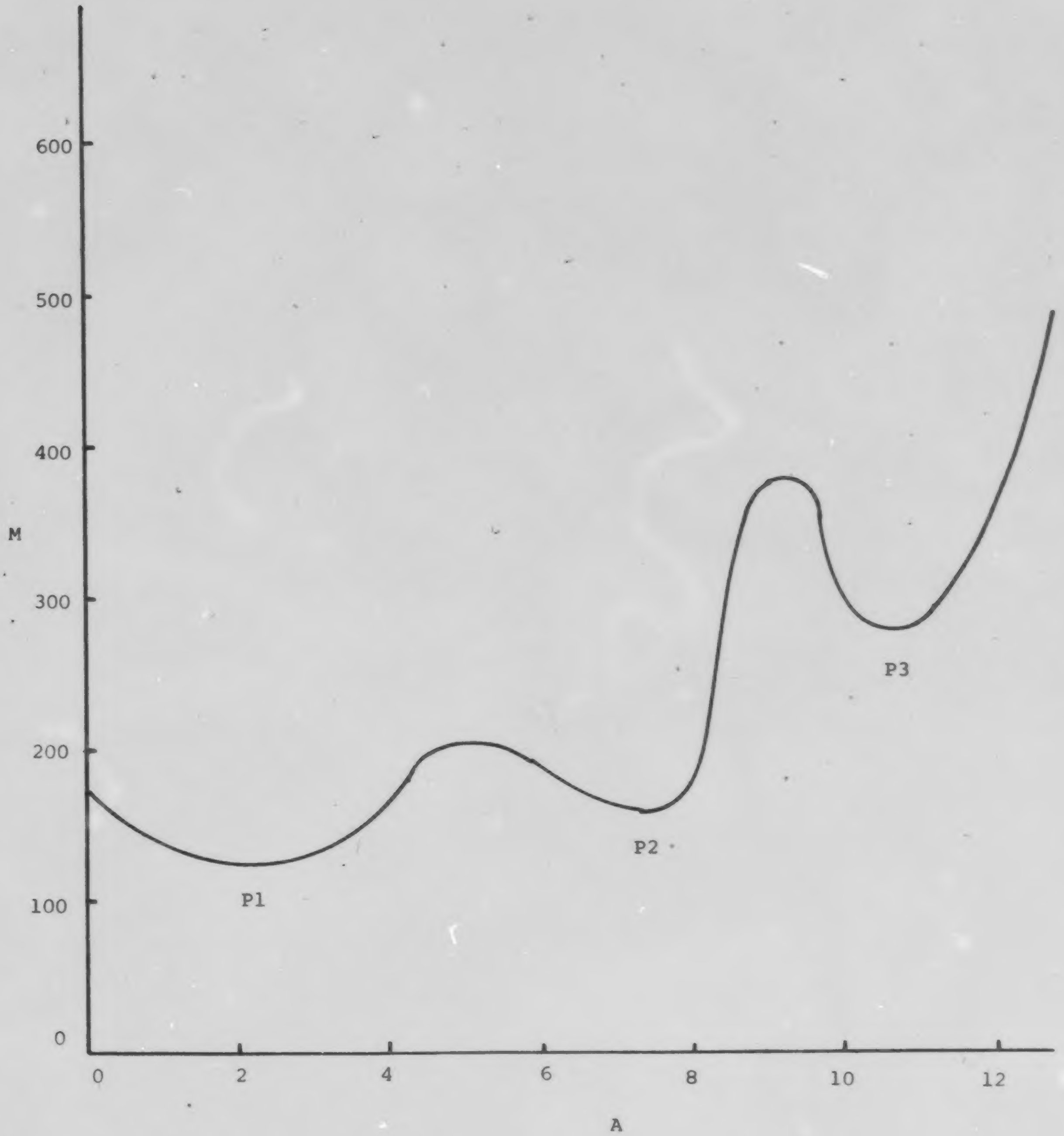


Fig. 4.1
A Minimization Curve

minimum points P_1, P_2, and P_3. P_1 is the true minimum point. The problem of determining the number of relative minimum points, their location, and the value of M at these points, is a very difficult problem. At the very least, it is more difficult than the original minimization problem. An effective and practical method of determining such points is still in the process of being developed. In actual practice the problem is resolved with the aid of the investigator's intuition, other knowledge of the phenomena the investigator may possess, or "brute force" computing. Thus, in our formulation of the problem we too will use the insight, etc. of the investigator.

The success of all minimization techniques rests on the assumption or hope that our model sufficiently well represents the phenomena under investigation and that our experimental data is sufficiently accurate, so that for the measure of closeness we select, there is only one minimum point. If there are local minimum points, we are further assuming that we know enough about the problem to "guesstimate" the initial value of the parameter so that our search routine will "find" the parameter value giving the minimum value of M. The implication of this statement can be more fully appreciated by examining Figure 4.1. We see that P_1 is the minimum point and that points P_2 and P_3 are local minimum points. Thus, we are hoping to guess a value of A in the neighborhood of $A = 2$, and to select an initial step size small enough so that we do not step over into the neighborhood of $A = 7$, or $A = 11$. In this example a step size of one unit or a half unit would be satisfactory. There are techniques for determining an initial guess and they are discussed in other sections of this chapter.

The question of what measure of closeness, i.e. what norm or criteria function, to use is not a trivial question since, as we have seen, different norms may give different values for the parameters. The selection of the norm is largely at the investigator's discretion and such techniques as plotting, comparison, other means of experimental verification, etc. are used in the final determination of the parameters. It is frequently helpful to determine the parameters using two or more norms thus providing a check on the consistency of the model and the work. This statement is based upon the premise (not proven) that if the model and the data are consistent, the value of the parameter yielding the minimum M for one norm should be equal to or very close to the value of the parameter yielding the minimum value for a different norm. Thus, if two widely different parameter values are obtained under two different norms there is serious reason to question the investigation. In summary, the determination of parameters occurring in most complex models is still an art, not a science.

An Automated Iterative Procedure or Algorithm

It has probably occurred to the student that the method described in the preceding chapter for determining the parameters could be automated and that such a procedure would be very desirable. This may be done by writing a computer program that mimics the necessary decision-making process and then carries out all of the prescribed alterations of the initial parameter estimates. Such a program is called an *algorithm*. An algorithm is a specified process or set of instructions for achieving a prescribed goal. In contrast, the programs that were developed in the first two chapters are called models because they are analogous to the mathematical expressions (equations) which describe or govern a phenomena. The principle characteristic that distinguishes an algorithm from a model is that the objective of the algorithm is usually precisely specified; whereas the objective of a model is usually to obtain insight into the phenomena. An algorithm may be part of a model since the achievement of a definite goal, such as the determination of the value of a function, may be an integral part of a model. The subroutine which calculates a random number is an example of an algorithm. There are different algorithms to solve the same problem or classes of problems. The choice of which algorithm to use can depend upon such factors as (1) the kind of computing facility available, (2) the accuracy desired, (3) the amount of computing permitted, etc.

A great deal of effort has been devoted to attempts to derive successful minimization algorithms and to more clearly delineate the conditions or hypotheses which will insure the success of such procedures. This task is not easy because the conditions for their success must not be so restrictive that these conditions prevent the procedure from being applied to a significant problem. Nevertheless, mathematicians have been quite successful in deriving algorithms which do succeed under very broad and general conditions. There also exist algorithms for which no proof of their general success can be given, but which do succeed admirably in a large number of cases. Proven algorithms may fail because one or some of their necessary conditions have not been met by the problem whereas quite heuristically based algorithms may succeed because of something exceptional about the problem under investigation.

The trial and error technique that we have described is an example of a frequently successful technique that is also quite general in that it may be readily adapted to a large variety of problems with an expected high probability of success. However, there is no guarantee that the method will converge, i.e. reach an answer, or that if it does converge, that it has converged to the correct solution. For these reasons it is mandatory, not just prudent, that the investigator check his results.

A Minimization Algorithm

The problem is to develop an algorithm that will determine the value of A that will render the quantity M a minimum. This will be accomplished by writing a computer program that mimics an orderly human decision making process. One possible orderly process is the following. Guess at an initial value for A and evaluate M. Denote this value by A0 and let H be a small positive number. Increase A0 slightly to A0 + H and again evaluate M using the new value of A. If this latest value of M is less than the original value, we have improved our initial guess. Suppose this were the case. In this event we should again increase A slightly, and evaluate M. If this new value of M is less than the previous value, we repeat the process until we reach a value of A which no longer decreases M. We denote this value of A by A1. In the event that our initial increase of A0 to A0 + H had not resulted in a decreased value of M we would have "proceeded in the opposite direction" by decreasing A to A0 - H and then evaluating M. If this had resulted in a value of M less than the original value, we again would have decreased A to A0 - 2H; M would have again been evaluated, compared with the previous value, and the process repeated until we arrived at a value for A which would no longer have decreased M. Just as before, this value of A is denoted by A1. At this point the value of the step size, H, is reduced to say H/10, and the entire procedure repeated using the point A1 as a starting point. Once the new endpoint is reached, the value of the step size is again reduced and the entire process is repeated using the last determined end point as the new starting point. In this way, by repetition, the extreme value can be determined to as many significant figures as desired. The procedure is terminated when the step size is less than a predetermined value.

A program for the determination of G, the growth coefficient, in the constant environment model using the empirically determined population data of the United States from 1790 to 1960 is given in Figure 4.2. The corresponding flowchart is listed in Figures 4.2a and 4.2b. The criteria function has been denoted by M1, rather than M, and the student should be able to readily follow the program by comparing it with the flowchart. The flowchart does not depict the evaluation of M1 in detail because M1 is arbitrary and the minimization program is independent of the specific criteria function. By simply changing the calculation of the criteria function, the program may be used to determine the value of A which minimizes other criteria for closeness. The calculation of M1 is accomplished in statements 300 to 385, and in this example M1 was chosen to be the sum of the squares of the deviations. We emphasize that the program may not always be successful since its success may depend upon a "good" initial estimate of the parameter and on the fact that there exist no local minimum points between the initial estimate and the extreme value of the parameter. A guarantee of the success of a given search method would require a careful statement of the conditions necessary for success and a proof of the convergence of the search method under the specified conditions. Such a task is difficult at best and would require advanced mathematics and so will not be discussed here. In defense of our cavalier attitude, it should be pointed out that such convergence proofs rely on hypotheses which are usually very difficult to insure and hence, are of little or no use to the investigator. In essence, one just tries and hopes for the best.

```
  1  REM     AUTOMATED ONE-DIMENSIONAL SEARCH
  5  REM     CONSTANT ENV. MODEL, U.S. POP. DATA, ONE YEAR TIME INTERVAL
 10  DIM P(50),E(50)
 12  DIM P1(100)
 15  GOSUB 500
 20  PRINT "INPUT THE INIT. GUESS A AND THE INIT. POP.  P(0)"
 25  INPUT A,P(0)
 26  PRINT
 30  PRINT "INPUT THE INIT. STEP SIZE  H  AND THE LIM. STEP SIZE  H1"
 35  INPUT H,H1
 36  PRINT
 40  PRINT "INPUT THE MAX. NO. OF ALLOWABLE STEPS,  C1"
 45  INPUT C1
 46  PRINT
 47  PRINT
 48  PRINT "THE VALUES OF  A,  M1  AND  H  ARE"
 50  LET C=0
 55  GOSUB 300
 60  LET M0=M1
 65  LET A=A+H
 70  GOSUB 300
 95  IF M1<=M0 GO TO 110
100  LET A=A-H
105  GO TO 200
110  LET C=C+1
115  IF C<C1 GO TO 125
120  GO TO 400
125  LET A=A+H:LET M0=M1
130  GOSUB 300
135  IF M1<=M0 GO TO 110
137  LET A=A-H
140  LET H=H/10
145  IF H<=H1 GO TO 450
150  GO TO 65
200  LET A=A-H
205  GOSUB 300
210  IF M1<=M0 GO TO 225
215  LET A=A+H
220  GO TO 140
225  LET C=C+1
230  IF C<=C1 GO TO 240
235  GO TO 400
240  LET M0=M1
245  GO TO 200
295  REM    INSTR.  NOS.  300  TO  385  EVALUATE  M1
```

Figure 4.2

```
300  LET P1(0)=P(0)
301  FOR I=1 TO 17
302  FOR J=0 TO 9
303  LET P1(J+1)=P1(J)+A*P1(J)
304  NEXT J
305  LET P(I)=P1(J+1):LET P1(0)=P(I)
306  PRINT P(I),E(I)
307  NEXT I
350  LET S=0
355  FOR I=0 TO 17
360  LET D=ABS(P(I)-E(I))
365  LET S=S+D*D
370  NEXT I
375  LET M1=S
380  PRINT A,M1,H
385  RETURN
400  PRINT "EXCEEDED MAX.  NO.  OF STEPS"
405  PRINT "THE VALUES OF  A  AND  M0  ARE"
410  PRINT A, M0
415  STOP
450  PRINT "SEARCH COMPLETE.  THE VALUES OF A,  M0  AND  C  ARE"
455  PRINT A, M0,C
460  STOP
495  REM     INSTR.  NOS.  500 TO  525  ENTER THE EXPERIMENTAL DATA
500  DATA 3.93,5.31,7.24,9.64,12.87,17.07,23.19,31.44
505  DATA 39.82,50.16,62.95,75.99,91.97,105.71
510  DATA 122.73,131.67,150.7,179.32
515  FOR J=0 TO 17
520  READ E(J)
525  NEXT J
530  RETURN
540  END

READY
```

Figure 4.2 (continued)

FLOW CHART OF MINIMIZATION ALGORITHM

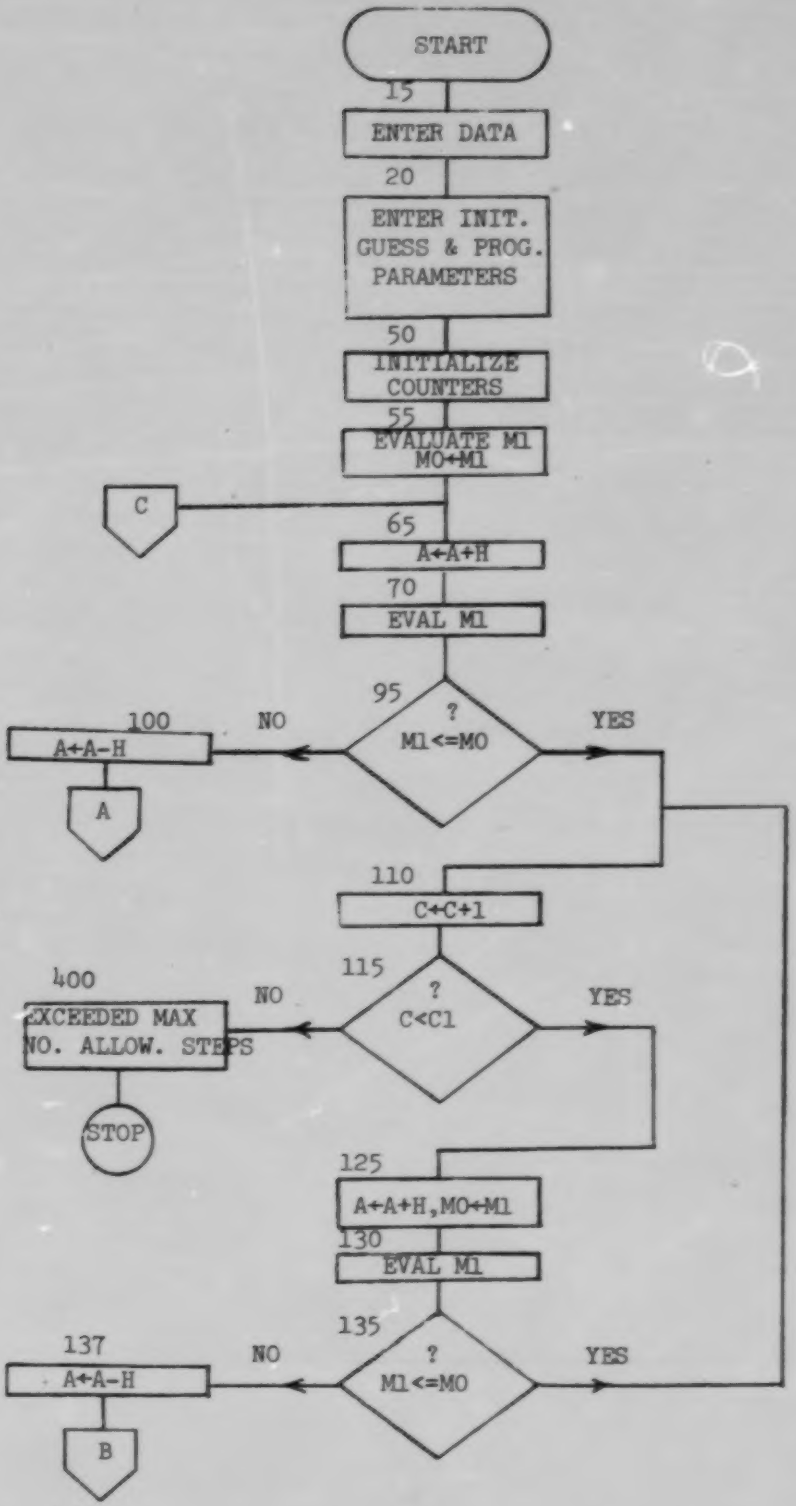

START
15
ENTER DATA
20
ENTER INIT. GUESS & PROG. PARAMETERS
50
INITIALIZE COUNTERS
55
EVALUATE M1 M0←M1
C
65
A←A+H
70
EVAL M1
95
? M1<=M0
NO
YES
100
A←A-H
A
110
C←C+1
115
? C<C1
NO
YES
400
EXCEEDED MAX NO. ALLOW. STEPS
STOP
125
A←A+H,M0←M1
130
EVAL M1
135
? M1<=M0
NO
YES
137
A←A-H
B

Fig. 4.2a

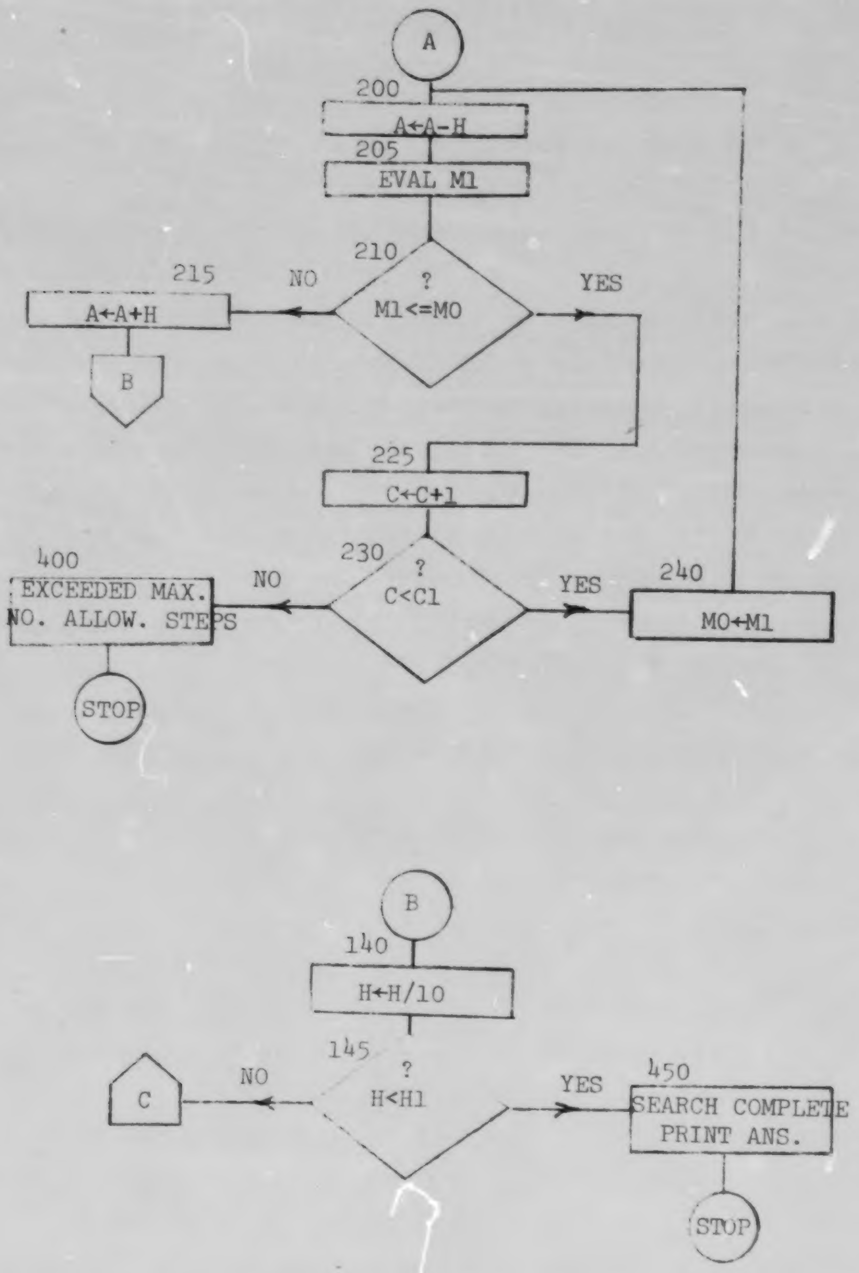

A
200
A←A-H
205
EVAL M1
210
?
M1<=M0
NO
YES
215
A←A+H
B
225
C←C+1
230
?
C<C1
NO
YES
400
EXCEEDED MAX.
NO. ALLOW. STEPS
STOP
240
M0←M1
B
140
H←H/10
145
?
H<H1
NO
YES
C
450
SEARCH COMPLETE
PRINT ANS.
STOP

Fig 4.2b

Statements 20 to 45 relate to the required input necessary to use the program. Instructions 20 and 25 are self evident. Statement 30 requires an initial step or increment size H, and H1 designates the smallest or stopping step size. In order to prevent possible endless looping or searching, a maximum number of allowable steps must be entered in the program. This is provided for in statements 40 and 45. For the results presented in this chapter, the value of 100 was used and most runs were completed in less than 20 to 25 steps.

Starting Values

In order to use the program developed in the previous section, it was assumed that the investigator had obtained, by one means or another, a good initial estimate of the parameter. In this section we indicate some methods for obtaining such estimates.

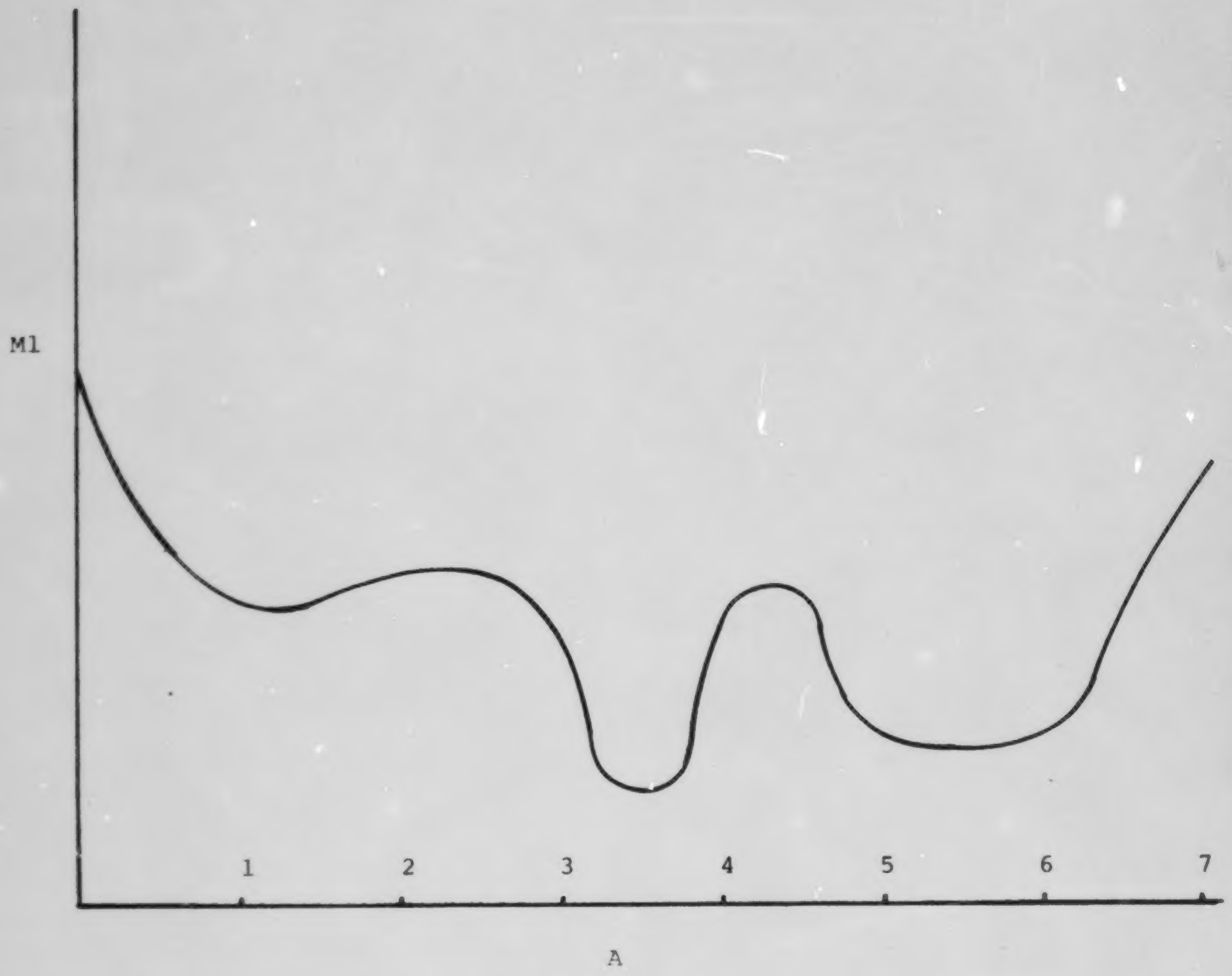
An obvious yet simple method for obtaining a starting value is to write a program for evaluating M1 for an arbitrary input value of the parameter. Then by examining the results for various values of the input parameter it is hoped that enough insight is gained to permit a close estimate of the required parameter value. This is the same procedure that was carried out in Chapter III to evaluate G. The process may seem simple and naive, but it is pleasantly quite successful. The method also has the valuable auxilliary attribute of providing further insight about the problem. The investigator may even find that his hypotheses or the program he developed using these hypotheses, is in error!

A second method is to guess at an interval which you "feel" contains the true value of the parameter. Denote the smaller of the interval end points by L and the larger of the interval end points by R. This interval is then subdivided into n equal intervals and M1 evaluated at each of the subdivision points. The subdivision point corresponding to the least value for M1 is then taken to be the starting point. The method can be used as a basis for an automated searching routine by merely applying the same subdivision process again to the subinterval containing the minimum value of M1 and repeating the process until the desired accuracy is obtained. (See problem 4). This process is called the uniform interval search method. The process will fail if the interval points are chosen incorrectly and Figure 4.3 depicts what could happen if the initial intervals were selected as shown. In this case, the uniform search method, using an interval size of one unit, would estimate that a value of A near 5 or 6 would be a "good" starting value whereas a value of A near 3 or 4 would be much better. In fact, a value A near 5 or 6 will result in the uniform interval search technique converging a value somewhere near $A = 5.5$ rather than in the neighborhood of $A = 3.5$.

A third method of determining an initial starting value is the random search method. This is just a simple variant of the uniform search method wherein a point in the interval from $A=L$ to $A=R$ is chosen at random and the value of M1 calculated. L and R denote the left and right extremities of the interval which is assumed to contain A. A second random point in the interval is then selected and M1 again evaluated. If the second value of M1 is less than that corresponding to the first random point, keep the second random point and repeat the whole process with a third randomly selected point. If the second value of M1 is not less than the first value of M1, keep the initial A value and repeat the whole process with a third random point. As many points are chosen as is desired or as one has time for. The process can fail in a manner similar to the way in which the equal interval search method fails and for the same reason. If it is known that the dependence of M on A is such that the curve of M vs. A is concave upward for all values of A in the interval, then it is possible to considerably shorten the random search technique. A curve is said to be concave upward in the interval (L,R) if, as A increases from L to R, the slope of the tangent line to the curve always increases. A concave upward curve "holds water" providing its minimum point is not at an end point. It also has the property that a line segment joining any two points on the curve is such that all points on this line segment lie above the curve, i.e. if the curve was "filled with water" all points on a line segment joining any two points on the water-curve interface would lie entirely within the water. It will be seen in the next section that if the criteria function is concave upward, this property is very useful.

Another Minimization Curve

Figure 4.3



Program Results

In this section we discuss the use of the program as well as present some implications of the results obtained from it. The program considers the Malthus model and is used to determine the growth coefficient corresponding to the U. S. population data as given in Table 3.1. The criteria for closeness was chosen to be the sum of the squares of the deviations and Table 4.4 lists typical program results. Runs numbered 1 to 15 correspond to a variation of the initial population and in the first run the initial population was chosen equal to that given in Table 3.2. For this run the value of 0.26175 was obtained and the value of the criteria for closeness, M1 was 4545.714. This parameter value was reached after eight evaluations from an initial guess of 0.262. In each run, the initial step size H was taken to be unit change in the last significant figure of the initial guess and the minimum step size was selected as one-thousandth of the initial step size. The runs were chosen in the order they occurred to your author at the time. Hence, they are not arranged in any completely systematic way nor are they selected in the most efficient manner. They are presented in this order to illustrate how one begins examining such results. Hindsight always beats foresight. The use of an excessive number of digits in presenting our results is done for purposes of illustration. The comments made in the previous chapter concerning the proper number of significant digits to use in an actual investigation still apply.

An examination of runs 2 through 15 of Table 4.4 reveals that as the initial population is increased, the value of M1 decreases. In fact, for an initial population of 11.0 million the value of M1 is 941.4217; almost one-fifth of that obtained in the first run. There is a corresponding large change in the calculated value of A which is 0.18101. The small value of M1 indicates that this latter curve is much closer (in the sense of least squares) to the actual population curve than is the curve corresponding to the first run. However, the difference between the initial population of 11.0 million as compared to the actual initial population of 3.93 million is probably far too large to tolerate and consequently the growth parameter of 0.18101 is not acceptable. This is an example of the well known fact that the least squares closeness criteria tends to force the

148

LEAST SQUARES

Run No.	Init. A	P(0) (millions)	A	M1	C
1	0.262	3.93	.26175	4545.714	8
2	0.262	5.00	.24272	3234.605	8
3	0.24	6.00	.22838	2382.822	9
4	0.24	7.00	.21630	1784.66	9
5	0.21	8.00	.20586	1376.996	9
6	0.21	9.00	.19667	1118.389	10
7	0.20	10.00	.18844	980.1222	11
8	0.19	11.00	.18101	941.4212	3
9	0.18	12.00	.17422	986.7569	9
10	0.16	11.5	.17754	954.4008	13
11	0.17	11.2	.17960	944.192	5
12	0.18	10.5	.18464	949.3965	14
13	0.18	10.8	.18244	942.0019	10
14	0.18	10.9	.18172	941.2874	7
15	0.18	11.1	.18030	947.39[illegible]9	3
16	0.1	3.93	.26175	4545.714	17
17	0.9	3.93	.26175	4545.714	23
18	-0.5	3.93	.26175	4545.714	23
19	-1.	3.93	.26175	4545.714	27
20	-2.	3.93	.26175	4545.714	37
21	-3.	3.93	.26175	4545.714	47

Program Results from Automated Search Routine

TABLE 4.4

two curves closer together for large ordinate values at the expense of smaller ordinate values. In this example the relative error for the smallest ordinate value or population is about 173% whereas the relative error for the largest ordinate value is less than 5%. Thus, the two curves are much closer, in a relative error sense, in that region where the ordinate values are the largest. We again remind the student that the time increment in our model was chosen to be ten years to correspond to the time increment of Table 1. Thus, the yearly growth coefficient is approximately one-tenth of this or 0.018. Actually this is not quite correct since a yearly growth would be compounded and hence, the yearly growth coefficient cannot be obtained by simply dividing the decade growth rate by ten. (See problems 6 and 7).

Since the value of M1 was increased after increasing the initial population beyond 11.0 million to 12.0 million, no further increase of the initial population was attempted. The student should note that such a simple test is certainly not a proof that the criteria function will continue to increase as the initial population is increased. This is merely an assumption on the part of your author. In the absence of such a proof, "some" confidence in the assertion that further increases of the initial population would not result in a smaller value for M1 could be obtained by trying a few other values of A greater than 11.0 million and noting that the value of M1 is increased. By varying the initial population as well as the value of A, in effect, a two-parameter search was conducted. The parameters were A, the growth coefficient corresponding to a decade; and P(0), the initial population. In the next chapter we shall discuss such two or more parameter search routines.

In runs 16 through 21, the initial population was always chosen equal to that of the tabulated initial population. The initial guess at the growth coefficient was varied and an examination of the results reveals that the same value for the growth coefficient is always obtained. This insensitivity to an initial guess is a very desirable property for the closeness criteria to possess. Frequently, one is not so fortunate and widely different initial guesses produce widely different final values. The cycle count reveals that as the initial guess strays farther away from the true value there is a corresponding increase in the number of cycles necessary to reach the final value.

By increasing the value of the step size H, as the initial guess was chosen further away from 0.26175, the cycle count could have been considerably decreased. In all of these problems, H was chosen to be 0.1. The fact that the process converged to the same final value for such a wide range of initial guesses indicates that the graph of M1 versus A "holds water", i.e. is concave upward. In mathematical optimization theory such a property is called convexity. This is a very desirable property and nearly all search type algorithms that have been proven to be successful require this property. (See Cooper and Steinberg).

Tables 4.5 and 4.6 display results for different criteria of closeness. In Table 4.5 the maximum of the magnitude of the deviation was the closeness criteria and in Table 4.6 the criteria for closeness was the maximum magnitude of the relative error. For an initial population of 3.93 million corresponding to the actual population in 1790, a value of 0.26254 was obtained for the growth parameter using the maximum of the magnitude of the deviation for the closeness criteria. A value of 0.27280 was obtained for the same initial population using the maximum of the magnitude of the relative error. Each of these parameter values is very close to the value of 0.26175 obtained by using the least squares measure of closeness.

Table 4.5 also shows a similar result when the initial population is varied. However, such is not the case when the closeness criteria is the maximum of the relative error. For this criteria, the best result is obtained for an initial population of 5.0 million, not 11.0 million. This is due to the fact that a large difference between the initial populations results in a large initial relative error which is independent of the magnitude of the growth coefficient. In fact, as the assumed initial population grows without bound the initial relative error also grows without bound; that is, becomes as large as you care to make it. Further examination of Figure 3.7 reveals that there is a break or discontinuity in the slope of the empirical curve at about 1930. The presence of this discontinuity can manifest itself by making the determination of the parameter much more difficult. For a positive growth coefficient, the Malthus (model) results in a computed curve of constantly increasing slope and hence, the obtaining of G by a comparison with a curve having

TABLE 4.5

Max. ABS (Deviation)

Run	Init. A	P(0) (millions)	A	M_1	C
1	.17	11.0	.18220	9.9945	5
2	.17	10.0	.18949	11.741	10
3	.18	3.93	.26254	27.50347	19
4	.20	7.00	.21704	18.05097	12

TABLE 4.6

Max. ABS (Relative Error)

1	.20	3.93	.27280	0.325571	17
2	.25	7.00	.135	.7811705	
3	.25	4.00	.27116	.2177319	9
4	.25	5.00	.25119	.2722646	19
5	.25	4.5	.26028	.2809245	10
6	.25	3.5	.28361	.3608289	7
7	.29	3.2	.29192	.3886122	5

Comparison of Results

a broken slope will be difficult. If the experimental data is "scattered" or has anomalies which are believed due to experimental error, it may be advisable to "smooth out" the data. The study and application of smoothing techniques is a matter for advanced mathematical statistics courses and will not be considered here. These techniques are usually averaging techniques and a very common, but crude, technique is the French Curve technique. Your author has seen this used with occasional surprising success. The technique consists in visually free-handing a curve through the data with the aid of a French Curve. The free-hand curve is then used as the empirical curve and the model parameters obtained by comparison with it. The procedure is easy to apply and can be used as a possible check technique.

Interpolation

The parameter determination methods that we have been discussing require the computed and experimental points to be corresponding points. The student will recall that two sets of data points are said to be corresponding data points if the abscissae of pairs of points coincide. As stated in the previous chapter, the achievement of corresponding points may be impossible due to limitations of experimental technique. It is the purpose of this section to describe a simple method of obtaining corresponding experimental and computed points. There are other, far more elegant and sophisticated, interpolation methods and they are presented in courses in numerical analysis. The method presented here will be based upon the elementary notion of proportion, and is called linear interpolation.

Since the computational and experimental points are interlaced points (See Figure 4.7), we have a choice of interpolating the experimental data to obtain experimental data points at points corresponding to the calculated points or we have the choice of obtaining from the calculated data, interpolated points corresponding to the experimental data points. We choose the former and the student should realize that the interpolation procedure to be described is independent of such a choice.

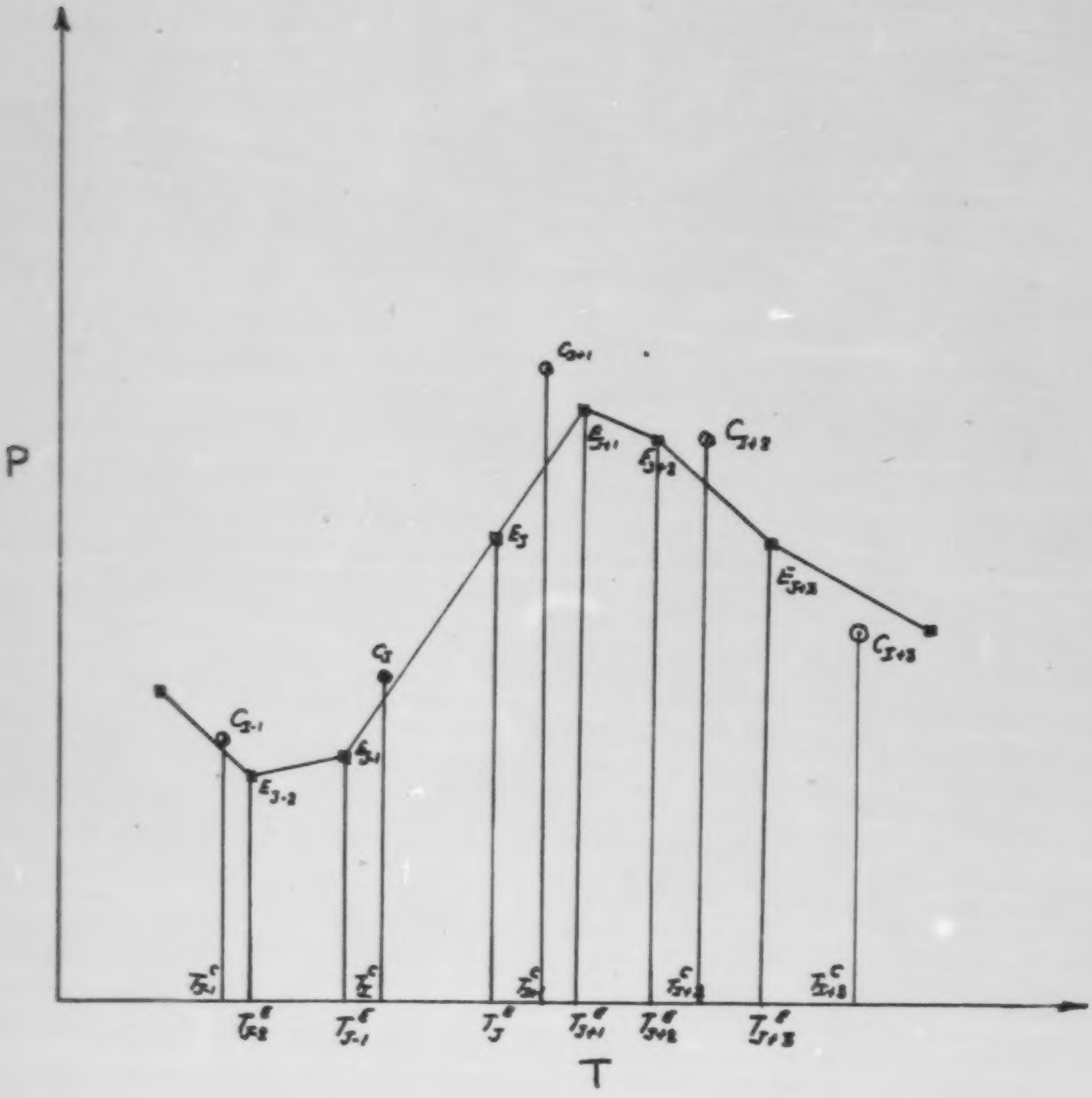

P
T
C_{J+1}
C_{J+2}
E_{J+1}
E_{J+2}
E_J
E_{J+3}
C_J
C_{J+3}
C_{J-1}
E_{J-1}
E_{J-2}
T^C_{J-1}
T^C_J
T^C_{J+1}
T^C_{J+2}
T^C_{J+3}
T^E_{J-2}
T^E_{J-1}
T^E_J
T^E_{J+1}
T^E_{J+2}
T^E_{J+3}

Fig. 4.7

In Figure 4.7 the points denoted by small circles represent calculated points and are equidistant on the T axis. The abscissae are indicated by T_I^C, I=1, 2, 3, ... The squares indicate experimental points and their abscissae, which are unequally spaced, are denoted by T_J^E, J=1, 2, 3, ... It is desired to obtain an estimate for the experimental data curve at the points whose abscissae are T_I^C, I=1, 2, 3, ... This can be done by determining two experimental points whose abscissae define an interval containing a T_I^C point and then estimating the value for the experimental curve at T_I^C. We will denote this estimated value by E_I. In the figure, the abscissae T_{J-1}^E and T_J^E of the experimental points E_{J-1} and E_J define an interval which contains the abscissae T_I^C of the calculated point C_I. A similar remark obtains about the abscissae of the experimental points E_{J+2} and E_{J+3} with regard to the abscissae of the computed point C_{I+2}. The estimation of the ordinate value of the point on the experimental curve whose abscissae is T_I^C will be based upon the assumption that the experimental curve is a straight line between the points E_{J-1} and E_J. This assumption implies that if the abscissae T_I^C were the midpoint of the interval (T_{J-1}^E, T_J^E) then the value of E_I would be the value of E_J plus one-half of the difference of the values of E_J and E_{J-1}. This is equivalent to saying that E_I would be the average value of E_J and E_{J-1} if T_I^C were the midpoint of the interval (T_{J-1}^E, T_J^E). If T_I^C is only one-fourth of the way from T_{J-1}^E to T_J^E, then the value of E_I is the value of E_{J-1} plus one-fourth of the difference between E_J and E_{J-1}. Thus, the important relation is the fraction of the distance $(T_J^E - T_{J-1}^E)$ that T_I^C is from T_{J-1}^E, and this fraction is merely the ratio $(T_I^C - T_{J-1}^E)/(T_{J+1}^E - T_J^E)$. We thus conclude that the value of the experimental curve for any point T_I^C lying in the interval (T_{J-1}, T_J) is given by

$$E_I = E_{J-1} + (E_J - E_{J-1})\ (T_I^C - T_{J-1}^E)\ /\ (T_J^E - T_{J-1}^E).$$

The same expression can be derived by using similar triangles and this is commonly done.

We reiterate, the method assumes that the experimental curve is a straight line (or is very close to being a straight line) between any two experimental points whose abscissae are adjacent points, so that the experimental value obtained by linear interpolation is equal to or close to the actual experimental value. This assumption is actually quite good if the points are close enough together and the data is not scattered or noisy. Because of experimental difficulties it is frequently the case that there is considerable scatter or noise in the experimentally determined data and for this reason it may be better to interpolate on the calculated data since they are usually not so scattered. The term scattered or noisy data refers to the location of the experimental points relative to a "reasonably smooth" and "simple" curve drawn through them. If the distances of the experimental points from this curve are random and widely fluctuating the scatter is said to be large whereas if the simple smooth curve passes through or is very close to each experimental point the scatter is said to be small. Your author realizes that this is a very heuristic definition and that it contains fuzzy terms, but he hopes that you, the student, get the general idea. The phenomena of scattering is usually dealt with by smoothing procedures which, as we stated previously, are best left to another course.

A computer subroutine to carry out the interpolation will consist of two parts. The first part will be a search routine to determine the pairs of experimental points whose abscissae contain the abscissae of a computed point. The second part will consist of the interpolation calculation. The required input will be the coordinates of the set of experimental points and the abscissae of the computed points. The set of experimental points should be arranged and labeled in order of increasing abscissae for ease of determining the ordinate pairs. If this is not the case, an ordering routine would be first used to label the set of experimental points each in the order of increasing abscissae. It is assumed that, as a result of calculation, the abscissae of the calculated points are so ordered. The flowchart in Figure 4.8, depicts the procedure and assumes that the labels I and J are such that increasing integral values of I and J correspond to increasing values of the abscissae of the respective computed and experimental points. Since the interpolation subroutine

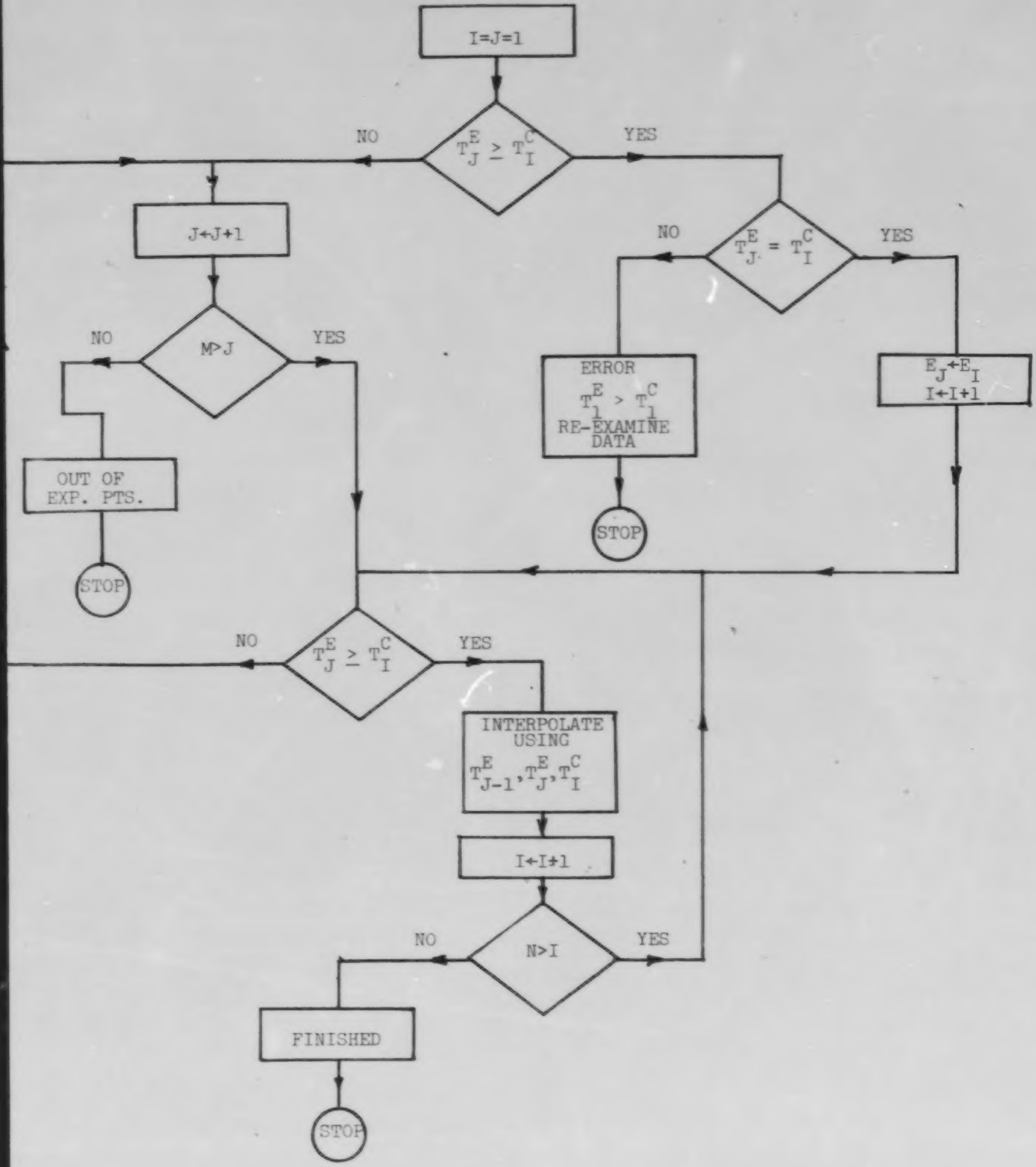

I=J=1
$T_J^E \geq T_I^C$
NO
YES
$T_J^E = T_I^C$
NO
YES
J←J+1
M>J
NO
YES
OUT OF EXP. PTS.
STOP
ERROR $T_1^E > T_1^C$ RE-EXAMINE DATA
STOP
$E_J \leftarrow E_I$
I←I+1
$T_J^E \geq T_I^C$
NO
YES
INTERPOLATE USING T_{J-1}^E, T_J^E, T_I^C
I←I+1
N>I
NO
YES
FINISHED
STOP

Fig. 4.8 157

requires the existence of both the experimental and the calculated set of points, it is assumed that these data sets are available upon entry to the subroutine. The subroutine is called whenever the continuance of the main program requires interpolated values. Consequently, the subroutine may be used many times in a single program. The flowchart was constructed assuming that

$T_1^E \leq T_1^C$ and that $T_M^E \geq T_N^C$ where M and N are the number of experimental and computed points respectively. Taken together, these two inequalities insure that the span of the abscissae of the computed points lies entirely within the span of the abscissae of the experimental points. The process of determining a function at points between where the function is defined is called interpolation and the process of determining the function at points outside the span of the abscissae points where it is defined is called extrapolation. The methods for extending or extrapolating a function to points exterior to the domain over which it is known are very similar to interpolation methods. However, experience has proven that the results are much less reliable and extrapolation should be avoided if reasonably possible.

PROBLEMS

CHAPTER IV

1. Evaluate G by modifying the program given in Figure 4.2 to
   (a) Minimize the maximum magnitude of the deviations.
   (b) Minimize the maximum magnitude of the relative error.
   (c) Minimize the sum of the squares of the relative error.
2. By varying the initial population determine the value of G for M1 equal to
   (a) The sum of the squares of the deviations.
   (b) The sum of the squares of the relative error.
   (c) The maximum of the magnitude of the deviation.
   (d) The maximum of the magnitude of the relative error.
3. For each of the four norms listed in problem 2 above, modify and use the program to find the minimum of
   (a) $A^2 + A - 10$
   (b) $A^4 + A^2 - 100$

   What are the differences in the results when the different norms are used?
4. (a) Modify the program given in Figure 4.2 to use an initial guess calculated gy one application of the equal search method. Using the data as given in Table 3.2 calculate an initial estimate to the Malthusian growth parameter with L=0.05 and R=0.95 and n=20.
   (b) Write a uniform search program and make up your own examples. Discuss the results.
5. (a) Draw a flowchart of the random search method.
   (b) Write a computer program using the random search method to find the minimum of

$$A^2 + A - 10 \quad \text{and} \quad A^4 + A^2 - 100 .$$

6. Modify the program of Figure 4.2 to calculate the population yearly for 1800 to 1960. By comparing the calculated population each decade with the populations given in Table 1 determine the growth coefficient.

7. Repeat problem 6 using

   (a) Max. magnitude of deviation norm.

   (b) Max. relative error norm.

   (c) Sum of squares of relative error norm.

8. The following record of the world's population growth from 1800 A.D. to 1968 A.D. is listed in the table below. (Emmel, 1973).

Year	1800	1850	1900	1930	1950	1960	1968
Pop. (Billions)	0.85	1.1	1.5	2.0	2.5	3.0	3.5

Calculate the growth coefficient using the least squares closeness criteria and a time period of one year. Emmel estimates the population of the world to be 4.5 billion people in 1980 and 7.4 billion in the year 2000. Do you agree?

REFERENCE

CHAPTER IV

Cooper, L. and Steinberg, D., 1970. Introduction to Methods of Optimization. W. B. Saunders, Co., Philadelphia, PA.

CHAPTER V

MULTIVARIABLE SEARCH METHODS

Introduction

Chapters III and IV considered the development of single variable search techniques for minimizing a criteria function that depended upon a single parameter. The methods were described by using as an example the determination of the growth coefficient G in the Malthus model to obtain a correspondence with the United States census data of 1790 to 1960. The criteria function was the least squares criteria and the initial population in the model was held constant and equal to the actual population in 1790. Later, in another sequence of runs, the initial population was also varied and for these cases it was found that the numerical value of the criteria function could be considerably reduced and that the value of G also changed accordingly. Since both P(0) and G were simultaneously varied, the minimization of the criteria function was, in effect, accomplished with the aid of a two variable search.

In this chapter we will investigate the development of two or more variable search methods with particular application to parameter determination methods. As was the case with single variable search methods, the techniques have very wide applicability since a great many diverse problems can be recast as minimization problems. Consequently, by developing the programs in such a way that the evaluation of the criteria function is done in a subroutine the programs may be applied to a broad range of problems. By constructing this subroutine to correspond to the criteria function appropriate to the particular minimization problem, the search routine program can be used as a method of solution to the problem.

An Example

To illustrate the development of a two parameter search program, we will consider the problem of determining G and G1 in a finite resource model of the United States population growth. The parameters will be determined by comparing the model results to the United States population data from 1790 to 1960. The procedure will be a direct extension of the single parameter method and both G and G1 will be obtained using the sum of the squares of the deviation as the closeness criteria. Hence, we will modify the finite resource model computer program to calculate the sum of the squares of the differences of the corresponding points of experimental and calculated populations. Furthermore, the time period in the model will be assumed to be one decade to coincide with the time period in the census data. Again, the procedure will be very heuristic and will consist in using the program to evaluate the sum of the squares of the deviations of the corresponding computational and experimental points for different pairs of values of G and G1. The values of G and G1 resulting in the smallest sum will be assumed to be the true values. This is a pure trial and error procedure, and as in the case of the guessing method for the determination of a single parameter, we hope that our intuition will be increased as the number of trials increases. Furthermore, for this simple model we are also hoping that only a small number of guesses should be required.

The alterations necessary to the finite resource program are very straight forward and very similar to the modifications made to the Malthus model. The program is listed in figure 5.1 and it should be self-explanatory.

Starting Values and Results

In the finite resource model the term G*P(I), for small populations, is very much less than G. Thus, for small starting populations, it is expected that the model will predict Malthus or exponential-like growth during the initial growth period. Furthermore, an examination of the graph of the actual United States population data also reveals that the initial growth is exponential. These facts intimate that using the value of 0.26 for the initial guess of G may be a good starting value for this parameter. The obtaining of a reasonable first guess for the auxialliary growth coefficient G1 is more difficult. However, it will be recalled that the derivation of the finite resource model was based upon the assumption that G1 was very much smaller in magnitude than G. Hence, a value of 0.00026 or 0.0026 might be a

```
RBFR

1 REM      TWO VARIABLE FINITE RESOURCE HEURISTIC SEARCH
2 REM
3 REM         U.S. POP. DATA, ONE DECADE TIME PERIOD
4 REM
5 REM
6 DIM P(50),E(50)
7 PRINT "INPUT G,G1 AND P(0)"
10 INPUT G,G1,P(0)
11 PRINT
12 PRINT
14 REM
15 REM          LINES 20-50 ENTER U.S.POP DATA
16 REM
20 DATA 3.93,5.31,7.24,9.64,12.87,17.07,23.19,31.44,39.82,50.16
25 DATA 62.95,75.99,91.97,105.71,122.78,131.67,150.7,179.31
30 FOR J=0 TO 17
40 READ E(J)
50 NEXT J
57 REM
58 REM          LINES 60-80 CALCULATE THE POPULATIONS
59 REM
60 FOR I=0 TO 17
70 P(I+1)=P(I)+(G-G1*P(I))*P(I)
80 NEXT I
82 PRINT "DECADE       CALC. POP.      EXP. POP.          DEV.          REL.
 ERROR"
84 PRINT
85 REM
86 REM     LINES 90-150 CALC. SUM OF SQUARES, S, AND PRINT RESULTS
87 REM
90 LET S=0
100 FOR I=0 TO 17
110 LET D=P(I)-E(I)
120 LET S=S+D*D
130 LET F=D/E(I)
140 PRINT I,P(I),E(I),D,F
145 PRINT
150 NEXT I
155 PRINT
160 PRINT
165 PRINT "  G              G1           P(0)            S"
170 PRINT
175 PRINT G,G1,P(0),S
180 END
190 END

READY
```

Program to Calculate G and G1 in the Finite Resource Model

Fig. 5.1

good starting value for the second parameter. If the limiting population, or carrying capacity, C, of the United States were known, the relation G/Gl = C would provide an initial estimate for Gl. This discussion on the obtaining of starting values was specific to the finite resource model and quite heuristic in nature. The student must remember that if very accurate initial guesses could be made, we would effectively have the answers to our problem and then presumably would have no need to use a search program in the first place. Other techniques for obtaining starting values are discussed later in this chapter.

A possible check on the accuracy of the program results can be obtained by noting that if the finite resource model does produce early time results which agree closely with the tabulated data, it is to be expected that the value of the growth coefficient obtained from the model would closely approximate the value obtained in the Malthus model. The checking of computer results in this manner is an example of using the results of a simpler program, the Malthus program, to check the results of a more complicated program, the finite resource program. It is not expected that the results should be identical because the programs and the hypotheses upon which they were built are not identical; however, any expected similarities in their results should be exploited for purposes of checking the correctness of the programs.

Tables 5.2a, 5.2b and 5.2c list results obtained from the program. The detailed results of a typical run are shown in Table 5.2a wherein G and Gl were chosen equal to 0.345 and 0.0017 respectively and the initial population, P(0), was chosen as 3.93 million. The value obtained for S was 300.836. Column 2 lists the calculated population and column 3 lists the actual population corresponding to each decade. The deviation, or difference of these two populations for each decade, is listed in column 4 and the fifth column lists the relative error of this deviation. It is seen that the deviation of maximum amplitude occurs in the last decade whereas the relative error of largest amplitude occurs in the seventh decade. Thus, the deviation of maximum amplitude occurs where the population is the largest whereas the maximum amplitude of the relative error occurs at a middle decade and for a quite small population. This indicates that different closeness criteria may produce model curves which compare differently with the experimental data. The results listed in Table 5.2b also confirm this fact. Table 5.2b illustrates the results of several computer runs corresponding to different values of G and Gl. In each run the initial population was identical to that recorded in 1790. Columns 6 and 7 of this table, list the deviation of maximum amplitude and the relative error of maximum amplitude respectively. The numbers appearing in parenthesis in columns 6 and 7 indicate the decade in which the respective maximum deviation and the maximum relative error occurred. The student will note that small changes in G and Gl produce large changes in all three measures of closeness; that is, in the sums of the squares of the deviations, the maximum deviation and the maximum relative error. The various criteria functions are thus quite sensitive to the parameters G and Gl. A comparison of the finite resource model results with the Malthus model results in table 3.5 shows that the least value for the sum of the squares of the deviations was 4545 for the Malthus model and only 301 for the finite resource model. This indicates that the finite resource model gives results that more closely fit the growth of the United States population from 1790 to 1960 than does the Malthus model. It is to be emphasized that these results confirm this conclusion only when the sum of the squares of the deviations closeness criteria is used as a basis for the comparison of the two models. If a similar analysis is carried out to ascertain the relative merits of the two models using a different closeness criteria such as the maximum deviation or the maximum relative error, a different conclusion might be reached.

Table 5.2b further reveals that the least value for the sum of the squares of the deviations is obtained for G = 0.345 and Gl = 0.0017 In contrast, the minimum maximum deviation is obtained for a value of G = 0.33 and a value of Gl = 0.00145, whereas a minimum value for the maximum relative error criteria is obtained for G = 0.35 and Gl = 0.0018. All three of these sets of values for both the growth coefficient and the auxilliary growth coefficient are fairly close. Consequently, it is safe to conclude that a reasonable value for G is 0.34 with an error of approximately ±0.01. Similarly, a reasonable value for Gl is 0.0017 with an error of approximately ±0.0002.

```
RUN

RBFR

INPUT G, G1 AND P(0)
?0. 345,    0. 0017,    3. 93

DECADE       CALC.  POP.     EXP.  POP.        DEV.          REL.  ERROR

 0              3. 93          3. 93        0              0

 1              5. 25959       5. 31       -. 0504065     -9. 49274E-03

 2              7. 02713       7. 24       -. 212874      -. 0294025

 3              9. 36754       9. 64       -. 272463      -. 0282638

 4              12. 4502       12. 87      -. 41984       -. 0326216

 5              16. 482        17. 07      -. 588045      -. 034449

 6              21. 7064       23. 19      -1. 48358      -. 0639751

 7              28. 3941       31. 44      -3. 04585      -. 0968783

 8              36. 8195       39. 82      -3. 00046      -. 0753506

 9              47. 2176       50. 16      -2. 94238      -. 0586598

 10             59. 7175       62. 95      -3. 23245      -. 0513495

 11             74. 2576       75. 99      -1. 73241      -. 0227978

 12             90. 5023       91. 97      -1. 46765      -. 0159579

 13             107. 802       105. 71      2. 09151       . 0197853

 14             125. 237       122. 78      2. 45705       . 0200118

 15             141. 78        131. 67      10. 1105       . 0767866

 16             156. 522       150. 7       5. 82185      . 0386321

 17             168. 873       179. 31     -10. 4366      -. 058204


  G              G1            P(0)          S

 . 345          1. 70000E-03   3. 93         300. 836

READY
```

Detail Printout of Fin. Res. Model

Table 5.2a

Run	G	G1	P(0)	S	Max. Dev.	Max Rel Error
1	.265	0.001	3.93	12561.95	-50.81(17)	-.411(10)
2	0.30	0.001	3.93	1534.52	-17.61(12)	-.256(7)
3	0.33	0.001	3.93	5690.329	+44.75(16)	-.297(16)
4	0.33	0.002	3.93	3743.266	-41.94(17)	-.234(17)
5	.33	0.0015	3.93	419.7755	- 9.14(17)	-.153(7)
6	.33	.0016	3.93	686.98	-16.84(17)	-.158(7)
7	.33	.0013	3.93	892.0854	18.49(16)	-.144(7)
8	.33	.0014	3.93	461.3196	11.64(15)	-.149(7)
9	.33	.00145	3.93	397.3647	8.99(15)	-.151(7)
10	.34	.0016	3.93	315.8525	10.77(15)	-.114(7)
11	.35	.0018	3.93	325.1555	-13.67(17)	-.0793(7)
12	.35	.0017	3.93	450.7575	14.56(15)	.111(15)
13	.35	.0016	3.93	941.342	20.19(15)	.153(15)
14	.345	.0017	3.93	301.0446	-10.45(17)	-.0969(7)
15	.345	.00165	3.93	356.3005	12.74(15)	.0967(15)
16	.3425	.00165	3.93	303.5757	10.46(15)	-.106(7)

Comparison of Results, Fin. Resource Model, U.S. Pop.
(Same Initial Population)

Table 5.2b

Run	G	G1	P(0)	S	Max. Dev.	Rel. Err.
1	.3	.0015	5.0	1619.42	-27.95(17)	.272(0)
2	.3	.0012	5.0	282.	7.91(15)	.272(0)
3	.33	.0014	5.0	2361.	27.61(15)	.272(0)
4	.33	.0015	5.0	1222.	21.08(15)	.272(0)
5	.33	.0017	5.0	422	-14.97(17)	.272(0)
6	.33	.0016	5.0	612	15.03(15)	.272(0)
7	.25	.0015	8.0	6641	-54.76(17)	1.035(0)
8	.25	.001	8.0	735	-18.38(17)	1.035(0)
9	.25	.009	8.0	366	- 8.57(17)	1.035(0)
10	.25	.008	8.0	506	11.52(16)	1.035(0)
11	.28	.001	8.0	5212	34.20(15)	1.035(0)
12	.28	.0012	8.0	1737	18.57(15)	1.035(0)

Comparison of Results, Fin. Resource Model, U.S. Pop.
(Different Initial Populations)

Table 5.2c

Table 5.2c lists results for different pairs of values of G and G1 corresponding to different initial populations. Using only these results as a basis we conclude that a variety of values may be obtained for the two growth coefficients by varying both the closeness criteria and the initial population. The student will recall that varying the initial population as well as the criteria function in the Malthus model, also resulted in different values for the growth coefficient. We again emphasize that the decision concerning the final choice of values for G and G1 must remain with the investigator since he alone must decide which closeness criteria he will accept. Frequently, a graphical portrayal of the results corresponding to the various best fits is an excellent aid in this decision. In addition, outside constraints, for example, the maximum permitted drug tolerance in a drug dosage model, can be an overriding consideration in the choice of a criteria function. We remark, in passing, that by varying the initial population, as well as both growth coefficients, we are in effect conducting a three variable search.

General Comments

In the course of the trial and error effort required to obtain estimates for the two parameters, it was noted that small changes in G1 produced much larger changes in the value of the criteria function than did small changes in G. Thus, the criteria function was found to be more sensitive to G1 than to G. It is frequently the case that the criteria function is more sensitive to some parameters than to others. It is also the case that there can be a great range in the magnitudes of the final values of the parameters; that is, the extremal values of some of the parameters may be very much larger than the extremal values of other parameters. Both of these facts suggest that during the search process, different parameters may require different search step sizes, or equivalently, the same search step size may not be appropriate for all parameters. If the criteria function is found to be insensitive to large changes in a parameter, it is appropriate to use a large step size when "searching on this parameter"; on the other hand if the criteria function is sensitive to small changes in a given parameter, it is appropriate to use a small search step size when minimizing with respect to this variable. Thus, in the automated two variable search routine that we will describe, we will make provision to permit a different search step size for each variable. A measure of the sensitivity can be obtained by making some trial runs with the previously described program.

In the next section, the development of a two variable search routine is presented. Your author wants to emphasize that the routine presented is just one of many different possible search routines. The existence of different search routines is due to the fact that in multivariable parameter minimization problems there are several ways to determine the many different possible paths which can be used in the conduct of the search. For example, it is possible to vary only one parameter until no further improvement is obtained and to then vary another parameter until no further improvement is obtained. This process is repeated until the effect of all of the parameters has been examined. The procedure is then started all over with the first parameter and the entire process is repeated until there is no further improvement in the criteria function for any parameter variation. This is the technique described in the next section. A different search routine consists in varying each parameter in turn by a specified small amount and then proceeded to the point yielding the most improvement in the criteria function. Using this point, the process is again repeated and a new neighboring point is found. The process is terminated when no neighboring point results in an improvement of the criteria function and the point at which this occurs is taken to be the extremum point. If more accuracy is desired, a smaller step size is selected and the entire process is again repeated. There are many other search routines and they are more or less independent of the criteria for closeness.

Because of the inherent difficulties associated with multivariable search problems, it is wise to use at least two different routines to see if the same parameter values are obtained. If each routine produces a significantly different set of parameter values, the results should be regarded as highly suspect, to say the least. As in the single variable case, a principle cause of such a divergence of answers is the presence of local relative minimum points. Such points occur far more frequently in two or more parameter minimization problems than in single parameter problems.

The discussion of relative minimum points for criteria functions of more than one parameter is fascilitated by the introduction of the notation of A, B, C, etc. for the parameters and the letter M for the criteria function. For example, the parameters A and B may be identified with the parameters G and G1 in the finite resource model or they may be identified with P(0) and G in the Malthus model. Our discussion will be restricted to the consideration of two parameter problems. The student should have no difficulty extending the ideas to three or more variable problems. As stated above, it is known that there may exist two or more relative minimum points. In terms of our notation, this means that there can exist two or more distinct pairs of values for A and B, called critical points, for which M is a local minimum. Thus, there can exist two distinct values of A and of B, say (A1, B1) and (A2, B2) such that the value of M at (A1, B1) is less than the value of M at local or nearby points of (A1, B1) and also the value of M at the point (A2, B2) is less than the value of M at points in the near neighborhood of (A2, B2). The value of M at (A1, B1) need not be the same as the value of M at (A2, B2). The points (A1, B1) and (A2, B2) are called relative minimum points.

The existence of relative minimum points in multiparameter minimization problems is more serious than the existence of such points in single parameter problems. Some of the reasons for this are: (1) such points usually occur with much greater frequency in multiparameter problems, (2) the determination of the number and location of these points is more difficult because the dimension of the search space is higher, and (3) the fact that the sensitivity of each parameter to a given step size may vary tremendously with each parameter frequently results in local minimum points having an inordinate affect on the search path.

In the section entitled Minimization in Chapter IV, several comments were made concerning difficulties associated with the application of search programs to single parameter problems. These comments apply even more forcefully to multiparameter problems and the student can rest assured that the greater the number of parameters to be determined the greater will be the difficulty of their determination. This is reflected in a greater increase in conceptualization time, programming time, and especially computing time. This latter increase is due to the accelerated expansion of the search space as the number of variables increases. R. Bellman speaks of this as "the curse of dimensionality" and the difficulty is well illustrated in the text by Cooper and Steinberg. In fact, as the number of variables increases the increase in the required computing time is so rapid that it is rare that search problems involving more than five or so parameters are attempted. Vigorous research efforts are presently being devoted to this problem by computer scientists and applied mathematicians.

A Minimization Algorithm

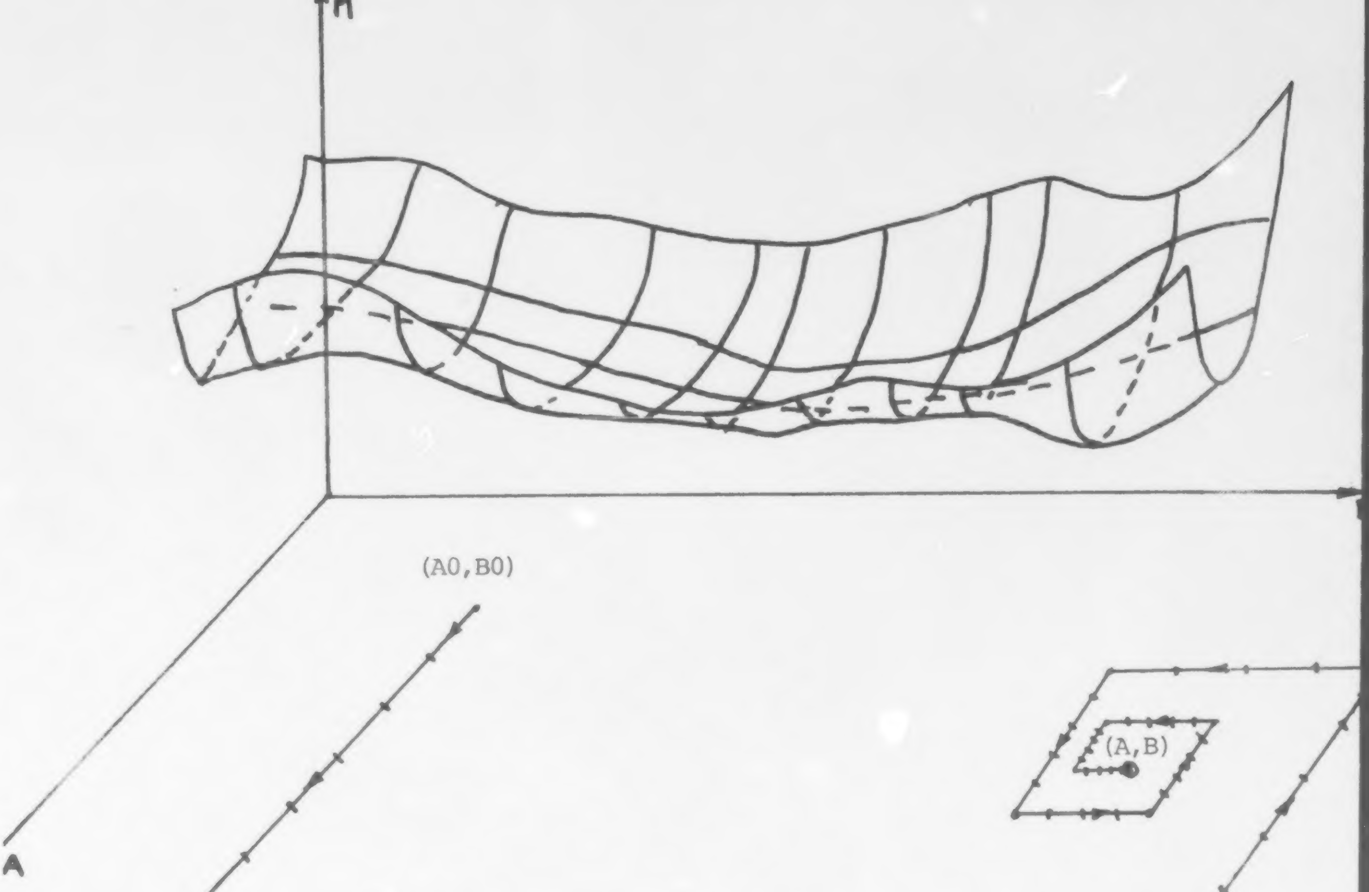
In this section we develop an algorithm to determine the values of the parameters A and B that will render the quantity M a minimum. This will be accomplished by writing a computer program that mimics an orderly human decision making process. One possible orderly process is the following. Select an initial value for A and B and evaluate M. Denote these values by A0 and B0 respectively, and let H and K be small positive numbers. H and K are the step sizes in the A and B directions respectively. Increase A0 slightly to A0+H and again evaluate M using the new value of A together with B0. If this value of M is less than the original value, we have improved our initial guess. Suppose this were the case. In this event we should increase A slightly to A0+2H, and again evaluate M using the original value B0 for B. If this new value of M is less than the previous value, we repeat the process until we reach a value of A, which when combined with B0 no longer decreases M. We denote this value of A by A1. In the event that our initial increase of A0 to A0+H had not resulted in a decreased value of M we would have "proceeded in the opposite direction" and decreased A to A0-H and again evaluated M using the value of B0 for B. If this had resulted in a value of M less than the original value, we would have further decreased A to A0-2H. M would have again been evaluated, compared with the previous value, and if it were less the process would have been repeated until we arrived at a value for A which when combined with B0 would no longer have decreased M. As above, this value of A is denoted by A1. We are now at the point (A1, B0) and, using the same procedures, will improve our guess on B0 by increasing B slightly from B0 to B0+K and comparing the values of M at the points (A1,B0) and (A1,B0+K). As before, if the value of M corresponding to the point (A1, B0+K) is less, we repeat the process until M is no longer decreased. This new point is denoted by (A1, B1). If the point (A1, B0+K) had resulted in an increased M, as compared to the value of M obtained at (A1, B0), we would have decreased B from B0 to B0-K and evaluated M. If the value of M were less, we would further decrease B from B0-K to B0-2K, compare the respective values of M, and repeat the process until M is no longer decreased. We denote this point by (A1, B1). This point can be thought of as our new starting point, (A0, B0), and the entire process repeated arriving at a new ending point (A1, B1). Using the last ending point as a new starting point, the process is repeated again and again until there is no change in either A0 or B0. These values can be accepted as the desired values for A and B or more refined values can be obtained by decreasing the size of the search steps H and K, to H/10 and K/10, using the final values of A and B, A1 and B1, as the new starting points and then repeating the entire process. By a repetition of these procedures still more refined values for A and B may be obtained.

The minimization problem for two parameters can be likened to the problem of finding the point on the ground closest to the bottom part of a motionless irregularly shaped balloon which is floating above the flat ground. The student may visualize the search process as a process of walking in the (A, B) plane (the ground) along a path which is alternately parallel to the A axis and then parallel to the B axis until the desired point is located. Figure 5.3 depicts this process and the bottom portion of the balloon surface is the M surface.

The conditions necessary for the success of the search routine can be inferred by an examination of Figure 5.3. It is intuitively evident that the surface must be convex and concave upward, that is "hold water", in a region containing the minimum point. Now the vertical projection onto the AB plane of the portion of the M surface containing the minimum point and which holds water is a section of the AB plane which we will call the "successful search space". It is so called because no matter which point in the space is selected for an initial point, the search routine will succeed.

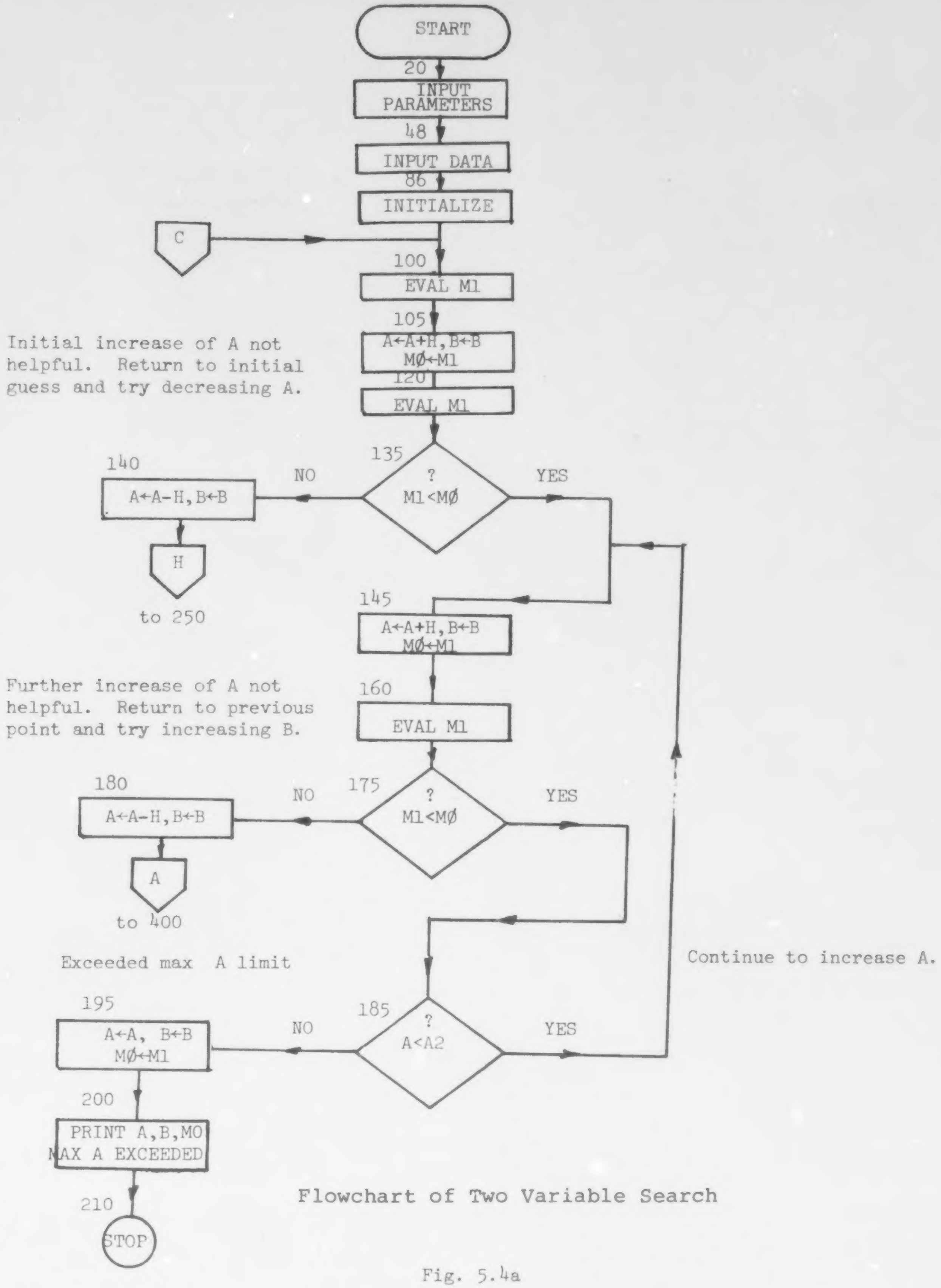
The actual determination of the successful search space is a task more difficult than the original problem of determining the minimum point. Consequently, we will not continue this discussion except to say that the techniques for easily delineating the successful search space are under intensive investigation.

A flowchart of the entire search routine is shown in Figures 5.4a to 5.4e. The comments appearing on the flowcharts are to clarify the procedure and the number appearing above the symbols refer to the corresponding statement numbers in the program listed in Figure 5.5. Figure 5.4a depicts the increasing of A and the possible consequences of such an action, and Figure 5.4b depicts the consequences of decreasing A. Figures 5.4c and 5.4d illustrate similar actions and the possible results for the second parameter B. Finally, Figure 5.4e illustrates the decisions necessary for changing the step size and for concluding the program. In this latter figure, the left-hand portion of the figure depicts the procedure necessary to reduce the search step sizes. There are two step sizes H and K, called the A step size and the B step size respectively, and both step sizes must simultaneously become smaller than their respective preassigned limiting step sizes E1 and E2. Since it is possible for one step


M
A
(A0,B0)
(A,B)
(A1,B0)
(A1,B1)

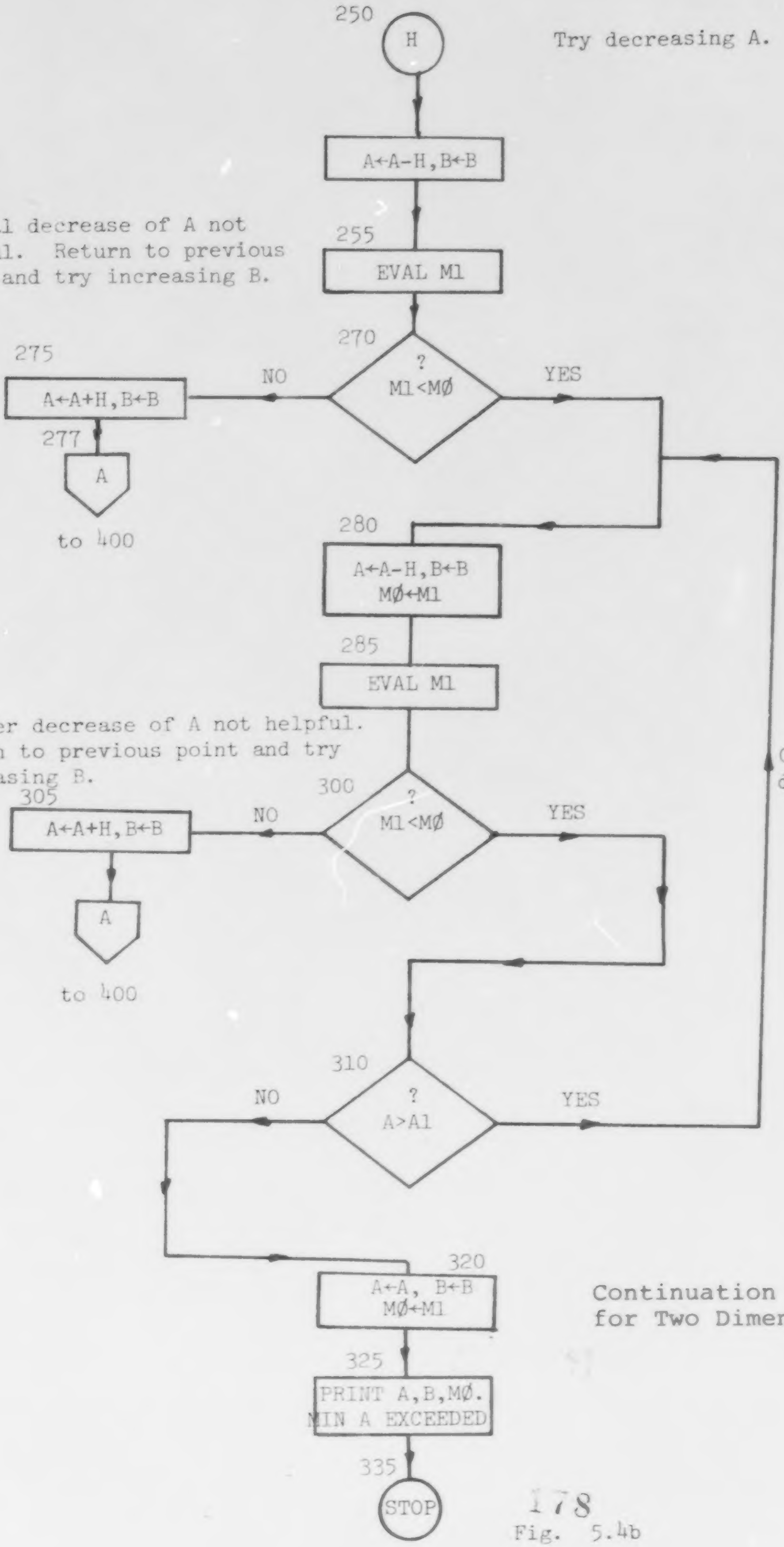
Two Dimensional Minimization Surface

Fig. 5.3


START
20
INPUT PARAMETERS
48
INPUT DATA
86
INITIALIZE
C
100
EVAL M1
105
A←A+H, B←B
M0←M1
120
EVAL M1
Initial increase of A not helpful. Return to initial guess and try decreasing A.
135
?
M1<M0
NO
YES
140
A←A-H, B←B
H
to 250
145
A←A+H, B←B
M0←M1
Further increase of A not helpful. Return to previous point and try increasing B.
160
EVAL M1
175
?
M1<M0
NO
YES
180
A←A-H, B←B
A
to 400
Exceeded max A limit
Continue to increase A.
185
?
A<A2
NO
YES
195
A←A, B←B
M0←M1
200
PRINT A,B,MO
MAX A EXCEEDED
210
STOP

Flowchart of Two Variable Search

Fig. 5.4a


250
H
A←A-H,B←B
255
EVAL M1
270
?
M1<M0
NO
YES
275
A←A+H,B←B
277
A
to 400
280
A←A-H,B←B
M0←M1
285
EVAL M1
300
?
M1<M0
NO
YES
305
A←A+H,B←B
A
to 400
310
?
A>A1
NO
YES
320
A←A, B←B
M0←M1
325
PRINT A,B,M0.
MIN A EXCEEDED
335
STOP

Try decreasing A.

Initial decrease of A not helpful. Return to previous point and try increasing B.

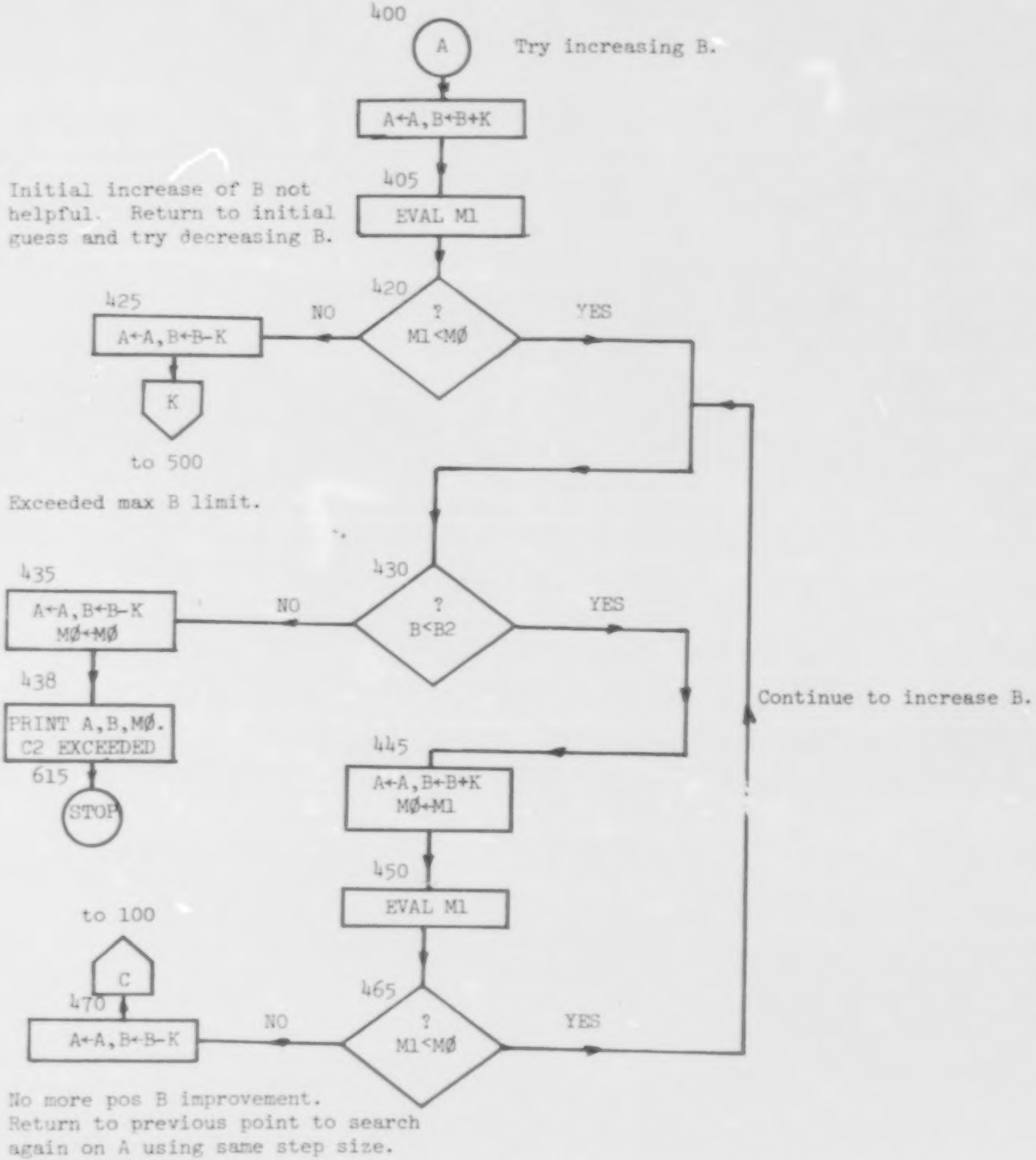
Further decrease of A not helpful. Return to previous point and try increasing B.

Continue to decrease A.

Continuation of Flowchart for Two Dimensional Search

178

Fig. 5.4b


400
A
A←A, B←B+K
405
EVAL M1
420
?
M1<MØ
NO
YES
425
A←A, B←B-K
K
to 500
430
?
B<B2
NO
YES
435
A←A, B←B-K
MØ←MØ
438
PRINT A,B,MØ.
C2 EXCEEDED
615
STOP
445
A←A, B←B+K
MØ←M1
450
EVAL M1
465
?
M1<MØ
NO
YES
470
A←A, B←B-K
C
to 100

Try increasing B.

Initial increase of B not helpful. Return to initial guess and try decreasing B.

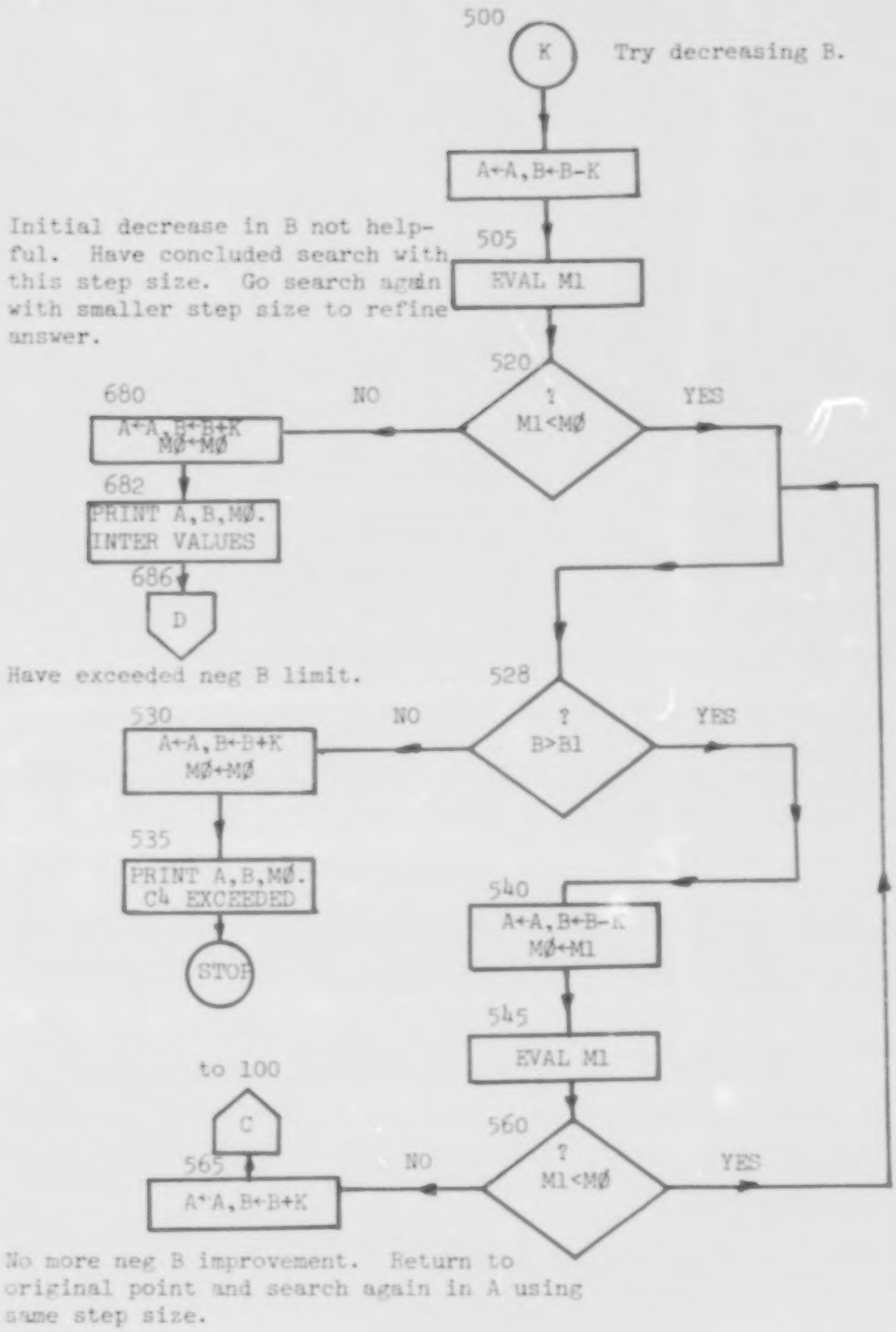
Exceeded max B limit.

Continue to increase B.

No more pos B improvement. Return to previous point to search again on A using same step size.

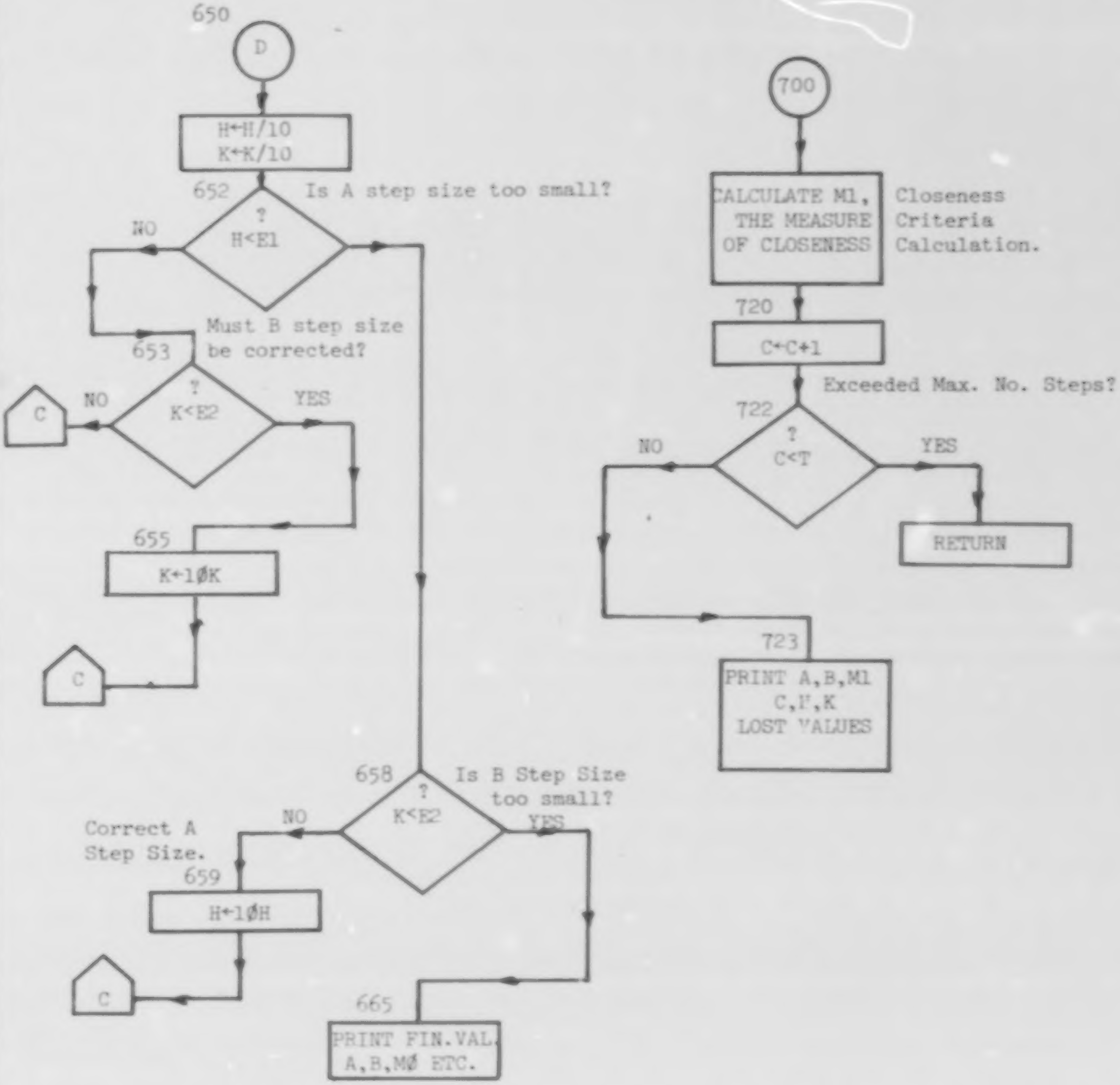
Continuation of Flowchart for Two Dimensional Search

Fig. 5.4c


500
K
Try decreasing B.
A←A, B←B−K
Initial decrease in B not helpful. Have concluded search with this step size. Go search again with smaller step size to refine answer.
505
EVAL M1
520
?
M1<M0
NO
YES
680
A←A, B←B+K
M0←M0
682
PRINT A,B,M0.
INTER VALUES
686
D
Have exceeded neg B limit.
528
?
B>B1
NO
YES
530
A←A, B←B+K
M0←M0
535
PRINT A,B,M0.
C4 EXCEEDED
STOP
540
A←A, B←B−K
M0←M1
545
EVAL M1
to 100
C
560
?
M1<M0
565
A←A, B←B+K
NO
YES
No more neg B improvement. Return to original point and search again in A using same step size.

Continuation of Flowchart
for Two Dimensional Search

Fig. 5.4d


650
D
H←H/10
K←K/10
652
Is A step size too small?
?
H<E1
NO
Must B step size
be corrected?
653
?
K<E2
NO
YES
C
655
K←10K
C
658
Is B Step Size
too small?
?
K<E2
NO
YES
Correct A
Step Size.
659
H←10H
C
665
PRINT FIN.VAL.
A,B,M0 ETC.
700
CALCULATE M1,
THE MEASURE
OF CLOSENESS
Closeness
Criteria
Calculation.
720
C←C+1
Exceeded Max. No. Steps?
722
?
C<T
NO
YES
RETURN
723
PRINT A,B,M1
C,H,K
LOST VALUES

Continuation of Flowchart
for Two Dimensional Search

Fig. 5.4e

size to be less than its limiting step size while the other step size is simultaneously larger than its limiting step size, provision must be made to recognize this fact and to act accordingly. Thus, instruction 652 determines if the A step size is less than the limiting step size of A and if it is not, a similar determination is made for the B step size in instruction 653. If the B step size is less than its limiting value, the B step size is multplied by 10 making it larger than its limiting step size. The search then proceeds with these step sizes. Analogous remarks can be made if the A step size is less than its limiting step size. This procedure insures that all search steps will be taken with step sizes not less than the preassigned limiting step sizes. The calculation of the closeness criteria and the determination of whether or not the preassigned maximum number of search steps has been exceeded is depicted in the right-hand portion of Figure 5.4e.

A program listing appears in Figure 5.5. We again remind the student that the program, as listed, is for the determination of the growth coefficient G and the auxiliary growth coefficient G1 in the finite resource model of the population growth of the United States from 1790 to 1960. The closeness criteria is the sum of the squares of the deviations between the experimental and the calculated populations. By suitably changing the subroutine for the evaluation of the criteria function, lines 700-716, other closeness criteria may be used. In addition, by suitably altering the experimental data statements, lines 48-80, together with the determination of the closeness criteria in lines 700-716, the program may be used to degermine the parameters occurring in two other parameter models. Thus, the program is quite flexible.

The program requires as input, an initial population P(0), an initial starting point (A,B); the respective step sizes in the A and B direction, H and K and the limiting or smallest permitted step sizes E1 and E2. There are two "safety checks" built into the program. The first check insures that the search path will not proceed without bound in either the A or the B direction. The second check insures that the search path does not loop back onto itself and thus result in an endless search, i.e. a loop.

The program assumes that the investigator is familiar enough with his problem so that he can initially specify a region in the AB plane in which the search should take place. It is assumed that this region is a rectangle whose sides are parallel to the A and B axis. The location of the sides of the rectangle must be specified as input and are accommodated in the program via statements 42 and 43. Statement 42 requires as input the minimum and maximum limiting search values in the A direction. These values are designated by A1 and A2 respectively. Statement 43 requires as input similar limiting values with respect to the B direction and they are denoted by B1 and B2. If any of these limits are exceeded by either an A or a B value in the search, the program stops and prints out the fact that the search path has "wandered out" of the permitted search region. The specifica-

```
1 REM  FINITE RESOURCE MODEL, LEAST SQUARES, U. S. POPULATION
5 PRINT
15 DIM P(50),E(50)
20 PRINT "A AND B ARE THE INITIAL GUESSES"
23 PRINT "P(0) IS THE INITIAL POPULATION"
24 PRINT "INPUT A, B, P(0)"
25 INPUT A,B,P(0)
26 PRINT
27 PRINT
30 PRINT "H  AND  K  ARE THE INITIAL STEP SIZES"
31 PRINT "E1  AND  E2  ARE THE LIMITING STEP SIZES"
32 PRINT "INPUT  H,  K,  E1,  E2"
33 INPUT H,K,E1,E2
34 PRINT
37 PRINT "T IS THE MAXIMUM ALLOWABLE NO. OF SEARCH STEPS"
38 PRINT "INPUT T"
39 PRINT
40 INPUT T
42 PRINT "A1  AND  A2  ARE THE MIN. AND MAX. PTS. OF  A  INTERVAL"
43 PRINT "B1  AND  B2  ARE THE MIN  AND MAX. PTS. OF  B  INTERVAL"
44 PRINT "INPUT  A1,  A2,  B1  AND  B2"
45 PRINT
46 INPUT A1,A2,B1,B2
48 REM INST. NOS. 50 TO 80 ARE DATA INPUT TO FIN. RES. MODEL
50 DATA 3.93,5.31,7.24,9.64,12.87,17.07,23.19
52 DATA 31.44,39.82,50.16,62.95,75.99,91.97
54 DATA 105.71,122.78,131.67,150.7,179.32
60 FOR J=0 TO 17
70 READ E(J)
80 NEXT J
84 REM INITIALIZING
86 LET C=0
95 PRINT "THE VALUES OF  M0,  A,  B  AND  C  ARE"
100 GOSUB 700
105 LET M0=M1
110 LET A=A+H\LET B=B
120 GOSUB 700
135 IF M1<M0GO TO 145
140 LET A=A-H\LET B=B
142 GO TO 250
145 LET A=A+H\LET B=B\LET M0=M1
160 GOSUB 700
175 IF M1<M0GO TO 185
180 LET A=A-H\LET B=B
182 GO TO 400
185 IF A<A2GO TO 145
195 LET A=A\LET B=B\LET M0=M1
200 PRINT A,B,M0
205 PRINT "EXCEEDED ALLOWED MAX. VALUE OF  A"
206 PRINT "THE VALUES OF  A,  B  AND  M0  ARE"
207 PRINT A,B,M0
208 STOP
210 STOP
250 LET A=A-H\LET B=B
255 GOSUB 700
270 IF M1<M0GO TO 280
275 LET A=A+H\LET B=B
```

Fig. 5.5

```
277 GO TO 400
280 LET A=A-H\LET B=B\LET M0=M1
285 GOSUB 700
300 IF M1<M0GO TO 310
305 LET A=A+H\LET B=B
307 GO TO 400
310 IF A>A1GO TO 280
320 LET A=A\LET B=B\LET M0=M1
325 PRINT A,B,M0
330 PRINT "EXCEEDED ALLOWED MIN  VALUE OF  A"
331 PRINT "THE VALUES OF  A,  B  AND  M0  ARE"
332 PRINT A,B,M0
335 STOP
350 PRINT
400 LET A=A\LET B=B+K
405 GOSUB 700
420 IF M1<M0GO TO 430
425 LET A=A\LET B=B-K
427 GO TO 500
430 IF B<B2GO TO 445
435 LET A=A\LET B=B\LET M0=M1
438 PRINT "EXCEEDED ALLOWED MAX  VALUE OF  B"
439 PRINT "THE VALUES OF  A,  B  AND  M0  ARE"
440 PRINT A,B,M0
444 STOP
445 LET A=A\LET B=B+K\LET M0=M1
450 GOSUB 700
465 IF M1<M0GO TO 430
470 LET A=A\LET B=B-K
472 GO TO 100
475 PRINT
500 LET A=A\LET B=B-K
505 GOSUB 700
520 IF M1<M0GO TO 528
525 GO TO 680
528 IF B>B1GO TO 540
530 LET A=A\LET B=B\LET M0=M1
535 PRINT "EXCEEDED ALLOWED MIN  VALUE OF  B"
536 PRINT "THE VALUES OF  A,  B  AND  M0  ARE"
537 PRINT A,B,M0
538 STOP
540 LET A=A\LET B=B-K\LET M0=M1
545 GOSUB 700
560 IF M1<M0GO TO 528
565 LET A=A\LET B=B+K
567 GO TO 100
570 PRINT
620 PRINT
645 PRINT
```

Fig. 5.5 (cont.)

```
650 LET H=H/10\LET K=K/10
652 IF H<E1GO TO 658
653 IF K<E2GO TO 655
654 GO TO 100
655 LET K=10*K
656 GO TO 100
658 IF K<E2GO TO 662
659 LET H=10*H
660 GO TO 100
662 PRINT
663 PRINT
664 PRINT
665 PRINT "THE FINAL VALUES OF  A,  B,  M0  AND  C  ARE"
667 PRINT A,B,M0,C
668 PRINT
669 PRINT "THE FINAL VALUES OF  H  AND  K  ARE"
670 PRINT H,K
675 STOP
677 PRINT
680 LET A=A\LET B=B+K\LET M0=M0
682 PRINT "THE INTERMEDIATE VALUES OF A, B, M0 ARE"
684 PRINT A,B,M0
685 PRINT
686 GO TO 650
689 PRINT
690 REM INST  NOS  700 TO 716 EVALUATE M
700 FOR I=0 TO 25
701 LET P(I+1)=P(I)+(A-B*P(I))*P(I)
702 IF P(I+1)<=0GO TO 705
703 NEXT I
704 GO TO 707
705 PRINT "THE POPULATION IS LESS THAN OR EQUAL TO ZERO"
706 STOP
707 LET S=0
708 FOR I=0 TO 17
710 LET D=P(I)-E(I)
712 LET S=S+D*D
714 NEXT I
715 LET M1=5
716 PRINT A,B,M1,C
719 REM INST  NOS  721 TO 730 PREVENT ENDLESS LOOPING
720 LET C=C+1
722 IF C<TGO TO 732
723 PRINT "THE LOST VALUES OF  A,  B,  M0  AND  C  ARE"
724 PRINT A,B,M1,C
727 PRINT "THE VALUES OF  H  AND  K  ARE"
728 PRINT H,K
729 PRINT "EXCESSIVE NUMBER OF STEPS"
730 STOP
732 RETURN
825 END
```

Fig 5.5 (cont.)

tion of the search region may be difficult; nevertheless, there is some validity to the statement that, "if the investigator knows so little about his problem that he cannot even vaguely specify the search domain, he certainly cannot specify reasonable starting values, and consequently, should not be attempting to solve the problem". Frequently, assistance in determining the boundaries of the search domain may be obtained by using the trial and error program previously described. In addition, it is occasionally possible to set the boundaries by using one of the routines for obtaining starting values. (See the section entitled Modifications).

Possible endless looping is prevented by counting the number of search steps and checking this number against the maximum number of allowable steps, C, which has been initially specified as input in statement 38. Since no search step may be taken without an evaluation of the criteria function, the search step count is accomplished by counting the number of times that the criteria function is evaluated. The counting and comparing is accomplished by statements 720 and 722. It is certainly true that in a well designed program this latter check should not be required. Nevertheless, your author, in recognition of Murphy's law, has developed the habit of including such a step counting routine in searching programs.

The student will recall that the purpose of the program was to estimate the value of the two growth coefficients G and G1 which occur in the finite resource model. However, the search routine was designed to be independent of a particular problem and so the search variables were designated as A and B. Consequently, in order to use the routine for the determination of the growth coefficients the fundamental finite resource model growth equation was written as

```
LET P(I+1) = P(I) + (A - B * P(I))*P(I)
```

and this is statement 701. The variable A is to be identified with the growth coefficient G and the variable B is to be identified with the auxilliary growth coefficient G1.

Modifications

The two parameter search routine just described constructs a search path that proceeds along a parameter until there is no further decrease in the criteria function and then the path is changed to proceed along the other parameter. A different search routine called the local search method, may be obtained by evaluating the criteria function at the three new neighboring points of the present point. The fourth neighboring point is the point from whence we have just come and by saving the value of the criteria function at this point we avoid the necessity of reevaluating the criteria function. The direction corresponding to the parameter step which resulted in the least value for the criteria function is the direction taken for the next step. The procedure is repeated using the new point as the starting point. If there is no further decrease in the criteria function in any direction, the step sizes are decreased and the process repeated using the last point as the new starting point. The procedure is stopped when a predesignated step size is reached.

It is possible to refine the method by using the magnitudes of the differences between the value of M at the point and its value at the neighboring points to provide a best direction in which to proceed. This direction will not necessarily be parallel to either of the parameter lines. For those students who have had the calculus this process is a finite difference, or secant equivalent, of a gradient technique for minimizing a function of two variables. This is frequently a more efficient method, especially if the number of parameters is large. Techniques for speeding up the search are known. However, since our purpose is to introduce the student to search methods and to illustrate their capabilities and limitations we will not discuss such refinements. Some of the problems will hint of these.

A still different search routine may be obtained by merely reversing the order of the variation of the parameters. Thus, the program is modified to first search on B and to then search on A. This simple variation can serve as an effective check on the accuracy of the first routine. In particular, your author has found that the simple reverse order search has proved very useful in ascertaining whether or not the former search has settled on a local minimum point.

Starting Values

The obtainment of good starting values for multiparameter search routines is necessary for their success. Such values may be obtained from several sources. The heuristic search program described at the beginning of this chapter may be the easiest way to obtain a "feel" for the situation and to thus permit a guess at reasonable starting values. Another way to determine starting values is to place a course grid over the AB plane and to then evaluate the criteria function at each of the grid points. The grid point corresponding to the minimum value of the criteria function is then used as the starting point. By applying this method again and again, each time with a smaller grid size and each time using the last determined extremal point as the origin of the grid for the next finer grid evaluation, the method can be used to determine the parameters. This is the two-dimensional equivalent of the uniform interval search method. By making the successive grid sizes smaller and smaller, the desired accuracy can be obtained. The existence of a local minimum point near a grid point can result in a great deal of computational effort to accurately determine the location of the local minimum point. If the size of the grid mesh is small, and the initial search space large, the method requires a large amount of computer time. A large mesh size reduces the calculational effort but increases the probability of "zeroing in" on a local minimum point. For these reasons the method must be used with care.

For some simple models a preliminary analysis of the BASIC equations may reveal relations among the variables which can be of assistance in the obtaining of starting values. The single population finite resource model is an example of such a model. In this model, preliminary analysis showed that the limiting population or carrying capacity, C, was related to the two growth parameters in a simple way. Thus for this model, if the empriical data indicates a limiting population, it is only necessary to estimate a starting value for one of the parameters, since the starting value for the remaining parameter is readily obtained from the equation.

Numerical Results

Figure 5.6 lists the detailed results obtained from a typical program run and a tabular listing of the results of several runs is presented in table 5.7. In all of these runs the initial population was chosen to be identical to the actual population and the runs were terminated when the search step size became less than the minimum permitted step size. For runs 9, 10 and 15 the minimum permitted step sizes for A and B were set at 0.000,01 and 0.000,000,1 respectively and for all of the other runs they were set at 0.000,1 and 0.000,01 respectively. Such excessive accuracy is unwarranted from the data; however, this accuracy was used for purposes of illustration and comparison of results. The results listed in runs 11 through 16 were obtained from reverse order searching, i.e. B was first varied and then A was varied. In all runs, the constraints on A and on B were each selected to be 0 and 2. The search path on the 16th run violated these constraints so the search was concluded before reaching terminal extremal values.

Most of these runs terminated in a small number of steps because the minimum permitted step size was fairly large. The number of steps necessary to arrive at a minimum point depends upon such factors as the accuracy of the starting values, the lengths of the initial and minimum permitted search step sizes, the criteria function, and the scatter in the experimental data. Some comments on the results are:

1. From the results of the first and second runs, as displayed in Table 5.7, it is seen that a 12% change in the initial value of B results in a 19% change in the value of the criteria function. If however, the minimum permitted step size in the second run is reduced by one-hundredth, the search yields extremal values almost equal to those obtained in the first run. Thus, it is seen that the final result can be very dependent upon the minimum permitted step sizes. A comparison of the results of run 3 with run 4 and the results of run 5 with run 6 shows that a 6% change in the initial value of A produces no significant change in the value of the criteria functions. An examination of the first six runs reveals that is is the initial value of B which is the more critical to the successful determination of the extremal values resulting in a minimum value of the criteria function.

```
A AND B ARE THE INITIAL GUESSES
P(0) IS THE INITIAL POPULATION
INPUT A, B, P(0)
?.34, .0017, 3.93


H  AND  K  ARE THE INITIAL STEP SIZES
E1  AND  E2  ARE THE LIMITING STEP SIZES
INPUT  H,  K,  E1,  E2
?.01, .001, .0001, .00001

T IS THE MAXIMUM ALLOWABLE NO. OF SEARCH STEPS
INPUT T
?100

A1  AND  A2  ARE THE MIN. AND MAX. PTS. OF  A  INTERVAL
B1  AND  B2  ARE THE MIN. AND MAX. PTS. OF  B  INTERVAL
INPUT A1, A2, B1 AND B2
?0, 2, 0, 2


THE VALUES OF  A,  B,  M0  AND  C  ARE
 .34             1.70000E-03   381.094        0
 .35             1.70000E-03   450.757        1
 .33             1.70000E-03   1199.07        2
 .34             2.70000E-03   9606.25        3
 .34             7.00000E-04   32604.3        4

THE INTERMEDIATE VALUES OF A, B, M0 ARE
 .34             1.70000E-03   381.094
 .34             1.70000E-03   381.094        5
 .341            1.70000E-03   347.069        6
 .342            1.70000E-03   321.989        7
 .343            1.70000E-03   305.917        8
 .344            1.70000E-03   298.915        9
 .345            1.70000E-03   301.044        10
 .344            1.80000E-03   461.952        11
 .344            1.60000E-03   442.398        12

THE INTERMEDIATE VALUES OF A, B, M0 ARE
 .344            1.70000E-03   298.915



THE FINAL VALUES OF  A,  B,  M0  AND  C  ARE
 .344            1.70000E-03   298.915        13

THE FINAL VALUES OF  H  AND  K  ARE

 1.00000E-04   1.00000E-05

STOP AT LINE 675
```

Results of a Typical Run of the Program Listed in Fig. 5.5

Fig. 5.6

RUN NO.	INITIAL VALUES A	B	H	K	FINAL VALUES A	B	M0	
1	.34	.0017	.01	.001	.344	.0017	298.9	1
2	.34	.0015	.01	.001	.333	.0015	357.7	1
3	.35	.0017	.01	.001	.344	.0017	298.9	1
4	.33	.0017	.01	.001	.344	.0017	298.9	1
5	.35	.0015	.01	.001	.333	.0015	357.7	1
6	.33	.0015	.01	.001	.333	.0015	357.7	1
7	.34	.0017	.1	.001	.339	.0016	309.3	2
8	.34	.0017	.1	.0001	.34	.001628	303.1	3
9	.6	.002	.01	.001	.3431	.001687	297.8	40
10	.1	.005	.01	.001	.3436	.001686	297.8	61
11	.34	.0017	.01	.001	.339	.0016	309.3	1
12	.34	.0015	.01	.001	.333	.0015	357.7	2
13	.35	.0017	.01	.001	.339	.0016	309.3	2
14	.33	.0017	.01	.001	.339	.0016	309.3	2
15	.6	.002	.01	.001	.3433	.001683	297.7	64
16	.1	.005	.01	.001	VIOLATED CONSTRAINT			

Summary of Results of Several Runs of Programs Listed in Fig. 5.5

Table 5.7

2. Runs 1, 7 and 8 indicate the effect of different step sizes.
3. Runs 9 and 10 are illustrative of the effectiveness of the program if the initial values of the parameters are quite different than the extremal values. In each of these runs the search path wandered considerably in the A-B plane and several steps were required to arrive at the final minimum value.
4. The results listed in runs 11 through 15, which were obtained by reversing the order of the search, show that virtually the same values are obtained and hence the order of the search is immaterial.
5. For those runs in which the indicated final value of the criteria function was considerably different than 298, it was found that a decrease of the minimum permitted step size would allow the search ton continue until a value of approximately 298. was obtained. This further illustrates the sensitivity of the extremal values to the minimum permitted step size.
6. In this example, the minimum value of the criteria function was usually obtained for each run. This was due to the smoothness and shape of the criteria function. If the maximum deviation had been chosen as the criteria function, such uniformly close agreement would not have been obtained.
7. There was considerable variation in the number of steps necessary to arrive at the extremal values. This variation increased as the magnitude of the minimum permitted step size decreased. Such behavior should not be unexpected since the traversing of even a relatively short distance will require many search steps if the magnitude of the search step size is sufficiently small.

The values obtained from a successful search must be checked to assure that a local extreme point has not be obtained. Some methods for checking the values are:

1. Rerun the program with different initial starting values for the search variables. If a different set of final values is obtained, the first set of variables may define a local minimum point.
2. Compare the values with other results obtained on similar problems.
3. Compare the values with any known empirical results.

4. Do the results seem "unreasonable"; that is, do they appear implausible, improbable, unrealistic, unexpected, etc.? If the answer to any of these criteria is yes, then the results are to be viewed as suspect. They may not be wrong; however, further checking and analysis is called for.
5. Conduct a reverse order search and compare the final results.

PROBLEMS

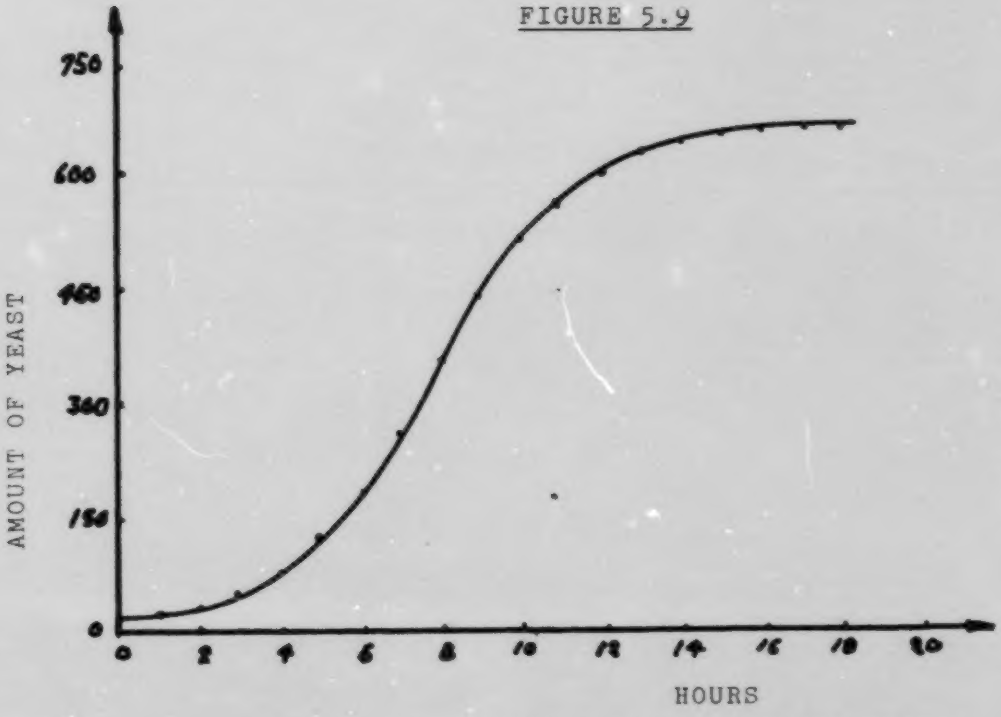
CHAPTER V

1. Pearl (1927) obtained a population growth curve for a yeast culture. It is given in Table 5.8, and is graphically displayed in Figure 5.9. The S or logistic shape of the curve is very evident and there is little scatter or noise in the data.
    - (a) Modify the program of figure 5.5 to determine the growth coefficients using one hour as the time period.
    - (b) Find the growth coefficients using for the closeness criteria
        - i. the maximum deviation
        - ii. the maximum relative error
        - iii. the sum of the squares of the relative error
    - (c) Find the growth coefficients assuming the time period is 15 minutes and the criteria function is the sum of the squares of the deviation.
    - (d) How do the growth coefficients of part (c) compare with those obtained in part (a)?
2. Modify the programs given in Figure 5.5 to calculate A and B using
    - (a) The maximum deviation.
    - (b) The maximum relative error.
    - (c) The sum of the squares of the relative error.
3. Find the yearly growth coefficients for
    - (a) 2a
    - (b) 2b
    - (c) 2c
4. For part (a) of problem 2 investigate the effect of
    - (a) The accuracy of starting values.
    - (b) The magnitude of initial search steps.
    - (c) The reversing of the order of the search.
    - (d) The magnitude of the minimum permitted search step.
5. Same as problem 3 only for part (b) of problem 2.
6. Same as problem 3 only for part (c) of problem 2.
7. Using the program given in Figure 5.5 and the maximum deviation as the closeness criteria, calculate the yearly growth coefficients. Compare them with the yearly growth coefficients obtained in 3a.

TABLE 5.8

Hours of Growth	Quantity of Yeast
0	9.6
1	18.3
2	29.0
3	47.2
4	71.1
5	119.1
6	174.6
7	257.3
8	350.7
9	441.0
10	513.0
11	559.7
12	594.8
13	629.4
14	640.8
15	651.1
16	655.9
17	659.6
18	661.8

FIGURE 5.9



YEAST CELL POPULATION GROWTH

8. Same as problem 7 using the maximum relative error as the measure of closeness. Compare to results of 3b.

9. Same as problem 7 using the sum of the squares of the relative errors as the increase of closeness. Compare to results of 3c.

REFERENCES

CHAPTER V

Cooper, L. and Steinberg, D. 1970. Introduction to Methods of Optimization. W. B. Saunders Co., Philadelphia, Pa, 19105.

Pearl, R. 1927. The Growth of Populations. Quarterly Review of Biology II, pp. 532-548.

Walsh, G. R. 1975. Methods of Optimization. J. Wiley & Sons, Inc., New York.

CHAPTER VI

LIFE TABLES

Introduction

The role of computer models in the management of natural resources is becoming more recognized. The principle task of a wildlife manager is the establishment of objectives and the determination of policies which will insure the achievement of those objectives. The simultaneous accomplishment of both of these tasks is very difficult since management objectives are often established as a result of compromise, and policies designed to achieve one set of goals usually preclude the achievement of another set of goals. As a result, the wildlife manager must be able to successfully analyze the expected outcome of implementing different policies prior to the actual enforcing of a prescribed policy. It is becoming increasingly apparent that computer simulation may be the tool which provides the wildlife manager with the capability of "pre-testing" his policies.

The management of wildlife populations subject to hunting harvest requires an estimation or prediction of the number of animals available for harvest. Such populations are usually distinguished by their sex as fawns or does, and by age classes as 1-year-old bucks, 2 to 5-year-old bucks, bucks over 6 years, etc. Consequently, the manager of a herd subject to harvest requires a model which will predict the time evolution of the population by sex specific age classes.

This chapter will consider the development of models whose objective is the prediction of the year by year evolution of the populations of the separate age classes in a population. The student will note that such models are based upon simple and direct extensions of previous work.

Development of the Model

The first model describing the time evolution of a population was the Malthus or exponential growth model. The fundamental equation of this model was

$$P(I+1) = P(I)+B*P(I)-M*P(I)$$

where B and M were the respective proportions of birth and deaths in a given period of time. These proportions were assumed to be constant over the entire time of evolution of the population. The equation expresses the fact that the population of the next generation is equal to the population of the preceding generation plus the number of births minus the number of deaths in that generation. It should be recalled that this Malthus model postulated an average birth rate and an average mortality rate and that, therefore, these rates are understood to be averages over all age classes in the population. The model thus ignored specific age class distinctions. In addition, the model made no allowance for differences in sex. The very young and the very old in a population are known to have a higher mortality rate than the middle aged and since it is further known that the birth rate varies with age; it seems reasonable to attempt to account for such effects in our model. This suggests dividing the population into age classes and then applying the fundamental law of change to each age group to enable the prediction of the population of each age class in successive years.

The classification of a population by age groups is called a Life Table. Since it is desired to describe the time evolution of an age class, an application of the fundamental law of change to the population in an age group gives

$$P(K+1) = P(K)+C(K).$$

If the time period is chosen to be one year, this equation states that the population of the K+1 year olds next year is equal to the population of the K year olds this year plus the change in population of the K year olds during this year. If the population of each age group is described in this manner, and if provision is made for estimating the newborn, it is seen that the initial life table population can be "upgraded" each year to produce a yearly succession of life tables which would describe the time history of the age groups in the population. With this background, the student should be able to readily follow the subsequent development.

For the description of herds consisting of large animals, such as deer, elk, bear, etc., which reproduce annually, it is convenient to choose the unit of time of the age group to be one year. Similarly, the unit of time during which the life table is upgraded will also be a year. Neither of these restrictions is essential. Since the fertility rates and the mortality rates of the separate age groups are sex dependent, it is realistic to further separate the populations into age groups by sex. We begin the development of the program by letting M(K) denote the male population whose age is K years and F(K) will denote the female population of the same age. D1(K) and D2(K) will denote the respective mortality rates of the male and female populations of the K^{th} age group.

We will also assume that there is no emigration nor immigration and consequently the sole cause of the change in population of an age group is mortality. Thus, an application of the above equation to each age group gives

M(K+1) = M(K)-D1(K)*M(K)

and

F(K+1) = F(K)-D2(K)*F(K)

for K = 1, 2, 3, ,,, where N denotes the number of age classes.

If the mortality rates are available in the memory of the computer, and if these equations are inserted into a loop with index K, it is evident that one sweep through the loop would produce the population of each age group for the next year. Another sweep through the loop would produce the population of the various age groups for the next successive year, etc. This process would enable us to calculate the time evolution of a life table. However, at the end of the calculation, the only populations available in the computer memory would be those contained in the life table of the last year of the calculation. Consequently, it would not be possible to examine the yearly change of a particular age group unless provision were made for printing out the life table each successive year. In addition, because past life tables are not available in the memory, it would be impossible to perform many useful calculations involving them. For these and other reasons, it is highly desirable to develop the program in such a way that all of the calculated information of interest is available at any time. Thus, we reformulate our program utilizing variables which have two subscripts; one subscript will specify the age group and the other subscript will specify the age of the life table. This different notation will permit the storing of intermediate results. The basic hypothesis will not be altered, just the program statements will be altered. The notion of altering the notation to fit the situation is a frequently occurring process when performing a mathematical analysis of a phenomena. In the latter event, it is often the case that a new mathematical notation is introduced or the existing mathematical notation is altered because it will enable the investigator to more readily analyze his problem. In so doing, the investigator does not change the basic biological or physical hypothesis or assumptions, rather he changes the mathematical notation to enable an easier analysis of the consequences of the basic hypotheses.

Consequently, we start anew to develop the computer program. Let M(K,I) denote the male population in the K^{th} age group after I time periods. I represents the age in time periods of the life table. Similarly, F(K,I) will denote the female population of the K^{th} age group in the life table whose age is I time periods. For example, if the length of each age group is a single year and the time period is also one year, M(6,4) denotes the number of males who are 6 years of age after 4 years. The statement F(7,5)=112 states that there are 112 7-year-old females alive after 5 years.

In this new notation, the fundamental age group calculations appear as

```
100 LET M(K+1,I+1) = M(K,I)-D1(K)*M(K,I)
```

and

```
110 LET F(K+1,I+1) = F(K,I)-D2(K)*F(K,I).
```

If K=8 and I=5, the first of these statements says that the number of 9-year-old males alive after 6 years is equal to the number of 8-year-old males alive after 5 years, minus the number of 8 year olds who die during the fifth year. This latter number is equal to the product of the mortality rate for 8-year-old males and the number of 8-year-old males alive after 5 years. Similar remarks apply to the female age group calculation. It is convenient to introduce the notation

S1(K)=1-D1(K) and S2(K)=1-D2(K).

The terms S1(K) and S2(K) can be interpreted as the survival rates of the K^{th} male and female age classes respectively.

Specifically, S1(K) and S2(K) are the proportion of males and females that were K years of age at the beginning of the I^{th} time period and who survived to become K+1 years of age at the beginning of the (I+1) time period. Since both D1(K) and D2(K) are mortality rates, they are each less than or equal to unity and hence, each of the survival rates is positive and less than or equal to unity, or zero. Ecologists frequently speak of the two sets of values, S1(K) and S2(K) as the set of age-specific survival rates. It is important to note that these rates are measured in units per length of time period. If the length of the time period is one year, they are proportions per year, i.e. they are the proportion of the population of the age class that survives each year. The survival rates have the dimension of reciprocal time units. It is not essential to assume that both the survival rates and the fecundity rates as well as the sex ratios of the newborn remain constant during the entire time of evolution of the population. Year to year variations in the weather, a variable food supply, size of the total population, the presence or absence of predators, etc. can all result in considerable yearly variations of the rates. The proper accounting for such variations is not easy because the obtaining of valid experimental data is difficult. Nevertheless, the inclusion in the model of variable fecundity and survival rates is readily accomplished with the aid of a subroutine or a stored array.

Statements 100 and 110 may be written in terms of the survival rates as

```
100 LET M(K+1,I+1) = S1(K)*M(K,I)
```

and

```
110 LET F(K+1,I+1) = S2(K)*F(K,I).
```

This is the form of the age group calculations that we shall use and is the form most frequently encountered in the literature.

The total number of offspring present at the beginning of the next time period is the sum of the newborn produced by each female age class during the present time period. It is known that the fecundity or reproductive rate varies with the age class and provision for such variation is made by denoting the rates by R2(K) where K is the index counting the age class. The number of newborn produced by the K^{th} female age group is equal to the number of females in that age group times the fecundity rate for that group, or F(K,I)*R2(K). The total number of newborn produced in a single period is

F(0,I)*R2(0)+F(1,I)*R2(1)+F(2,I)*R2(2)+......+F(N,I)*R2(N).

If we denote the number of newborn in the I^{th} period by C(I), a possible computer program for calculating C(I) is:

```
LET C(I)=0
FOR K=0 TO N
LET C(I)=C(I) + F(K,I)*R2(K)
NEXT K
```

Usually R2(0)=0, i.e. no first age group animals give birth. For ease of coding, it is sometimes convenient to carry the extra term. Since reproductive rates are the highest during the years immediately following the onset of adulthood, it follows that as K increases, R2(K) will increase rapidly to a maximum, level off, and then gradually decrease. The obtaining of such fertility data is quite difficult and it is known that fertility data established for a given herd in a specific habitat probably will not correspond to that established for the same herd existing in a different habitat. It is also assumed that there is always a sufficient number of males present in the population to service all of the females.

To obtain the number of newborn males and females in a single year, it is convenient to assume that the sex ratio of the newborn is independent of the age class. Hence, if F1 denotes the fraction of the newborn that are males, the number of newborn males and females respectively is given by

$$M(1,I+1) = F1*C(I)$$

and

$$F(1,I+1) = (1-F1)*C(I).$$

The determination of the mortality rates D1(K) and D2(K) may also be a difficult, time-consuming and expensive task. It is sometimes possible to use the model as a means of obtaining these values. This may be accomplished if an accurate set of life tables is available for a succession of years. An assumption of what are thought to be "reasonable" mortality rates is used as input and the program is run. If the computer results are in agreement with the available life table populations for a sequence of successive years, then it is reasonable to assume that the mortality rates are correct. If the program results do not agree with the tabulated data, another set of mortality rates is tried and results again compared. This trial and error process is quite inefficient, but the problem is difficult. Emlen (1973) presents a detailed discussion of the development of life table data. We also assume that both the survivability rates and the birth rates do not change from one time period to the next.

Leslie (1945), using matrix algebra, put forward a deterministic model for the prediction of life tables and hence, many ecology texts use matrix algebra when discussing life tables. Our use of BASIC provided a direct formulation of the problem in the programming language. It is interesting to note that if the matrix model of the problem were programmed for a computer using BASIC, the ability of the BASIC language to perform matrix operations would simplify the programming. However, since most of the matrix elements are zeroes, such a BASIC expression of the problem would be very wasteful of computer time. Our direct expression is much more economical of computational time, and as we shall see later, permits a ready modification of the program model.

Because the survival rates are all positive and less than unity, they also may be regarded as probabilities. Usher (1972) considers the sets S1(K) and S2(K) as transition probabilities from one time to another and in this way brings to bear many of the mathematical results of probability theory. All of these interpretations appear in the literature.

An Example

In the previous section, it was assumed that the sex ratio of the newborn was independent of the age class. For example, if 45% of the newborn from the third age class is male, it was assumed that 45% of the newborn from all of the other age classes were also male. The empirical data of Lowe (1969) for the red deer population on the Island of Rhum has been analyzed by Usher (1972) and resulted in fecundity rates which indicate a variation from age class to age class of the sex ratios of the newborn. Thus, it is necessary to modify the newborn calculation to account for this variation. It is assumed that, for each age class, the birth rate for male and female offspring is known. This data can be obtained from a knowledge of the newborn sex ratios from each age class. Denote the birth rates for the male and female offspring respectively by $Y1(K)$ and $Y2(K)$ where K is the index counting the age class. Because the fraction of newborn males plus the fraction of newborn females must equal the fraction of newborn of each age class, it is evident that

$$Y1(K) + Y2(K) = R2(K) \quad \text{for all } K.$$

Now, the number of newborn males for the K^{th} age class is $Y1(K)*F(K,I)$, and thus, the total number of newborn males is given by

$$M(1,I+1) = Y1(1)*F(1,I)+Y1(2)*F(2,I)+\ldots\ldots+Y1(N)*F(N,I).$$

The number of newborn females is given by a similar expression.

As an example, we consider the application of our life table model to the red deer population on the Island of Rhum using the data as presented by Usher (1972). Table 6.1 gives the survivability and fertility rate data by sex for each age class. Figure 6.2 lists the program and the following paragraph describes the program. Because of the simplicity of the program, a flowchart is not presented.

In order that the student may more easily follow the program, we again emphasize that the index K counts the age classes and the index I counts the time periods. In this program, both the time period and the time span of each age class are assumed to be one year. Statements 40 to 100 provide for the inputting of the survivability rates per age class. The fecundity rate data is entered in statements 110 to 165. By changing these rate data, other populations may be studied. For the initial population distribution, your author for no good reason, chose a population of 100 deer for each age class, both male and female. These starting populations are entered in lines 170 to 215. The actual calculation begins at line 405 and is continued through line 490. In this calculation, the index I counts the generations. At the start of each generation, lines 410 and 411 initialize the number of newborn counters, and the calculation of the number of newborn for each generation is accomplished by instructions 450 and 460. Lines 430 and 440 calculate the number of male and female deer in the next age group for the next period. The total population, P(I), is calculated each period and this is accomplished by instructions 472 to 475. Each year the number of newborn, as well as the total population, is printed out by statement 480. Finally, statements 510 to 530 print the population of each age class for the last time period.

Survivability and birth rates by Sex for Each Age Class

(From Usher, 1972)

AGE(K)	S1(K)	S2(K)	Y1(K)	Y2(K)
1	0.718	.863	0	0
2	.990	.902	0.202	.214
3	.990	.882	0.419	.444
4	.990	.879	0.434	.459
5	.990	.862	0.362	.589
6	.991	.840	0.363	.589
7	.734	.808	0.353	.576
8	.496	.507	0.376	.612
9	.370	.326	.422	.353
10	.848	.864	.417	.348
11	.821	.824	.464	.388
12	.781	.810	.464	.388
13	.720	.735	.464	.388
14	.611	.680	.464	.388
15	.364	.529	.464	.388
16	0	0	.464	.388

Fig. 6.1

```
RBLT1

5 REM    RHUM ISLAND DEER LIFE TABLE PROGRAM
6 REM
7 REM
20 DIM M(16,26),F(16,26),S1(16),S2(16),Y1(16),Y2(16)
39 REM
40 REM              THESE DATA ARE MALE SURVIVABIITY RATES
41 REM
45 DATA .718,.99,.99,.99,.99,.991,.734,.496
50 DATA .37,.848,.821,.781,.72,.611,.364,0
60 FOR J=1 TO 16
65 READ S1(J)
70 NEXT J
74 REM
75 REM              THESE DATA ARE FEMALE SURVIVABILITY RATES
76 REM
80 DATA .863,.902,.882,.879,.862,.84,.808,.507
85 DATA .326,.864,.825,.81,.735,.68,.529,0
90 FOR J=1 TO 16
95 READ S2(J)
100 NEXT J
109 REM
110 REM              THIS FECUNDITY DATA IS FOR MALE NEWBORN
111 REM
115 DATA 0,.202,.419,.434,.362,.363,.355,.376
120 DATA .422,.417,.464,.464,.464,.464,.464,.464
125 FOR J=1 TO 16
130 READ Y1(J)
135 NEXT J
139 REM
140 REM              THIS FECUNDITY DATA IS FOR FEMALE NEWBORN
141 REM
145 DATA 0,.214,.444,.459,.589,.589,.576,.612
150 DATA .353,.348,.388,.388,.388,.388,.388,.388
155 FOR J=1 TO 16
160 READ Y2(J)
165 NEXT J
169 REM
170 REM                INITIAL MALE POPULATION
171 REM
175 FOR K=1 TO 16
180 LET M(K,0)=100
185 NEXT K
199 REM
```

Fig. 6.2

```
200 REM                 INITIAL FEMALE POPULATION
201 REM
205 FOR K=1 TO 16
210 F(K,0)=100
215 NEXT K
230 PRINT "N=NO. OF YRS. EVOLUTION. N MUST NOT EXCEED 25"
235 INPUT N
240 PRINT
241 PRINT
242 PRINT
245 PRINT "                   PROGRAM RESULTS"
250 PRINT
255 PRINT
395 PRINT "PERIOD      NEWBORN MALES     NEWBORN FEM.    TOT. POP."
396 PRINT
399 REM
400 REM                   I COUNTS THE TIME PERIODS
401 REM
405 FOR I=0 TO N
410 M(1,I+1)=0
411 F(1,I+1)=0
414 REM
415 REM                   J COUNTS THE AGE CLASSES
416 REM
420 FOR K=1 TO 15
430 M(K+1,I+1)=S1(K)*M(K,I)
440 F(K+1,I+1)=S2(K)*F(K,I)
450 M(1,I+1)=M(1,I+1)+Y1(K)*F(K,I)
460 F(1,I+1)=F(1,I+1)+Y2(K)*F(K,I)
470 NEXT K
472 LET P=0
473 FOR S=1 TO 16
474 LET P=P+M(S,I+1)+F(S,I+1)
475 NEXT S
480 PRINT I+1,M(1,I+1),F(1,I+1),P
490 NEXT I
495 PRINT
496 PRINT
497 PRINT "POPULATION BY AGE CLASS FOR THE LAST TIME PERIOD"
500 PRINT "AGE CLASS, NO. OF MALES, NO. OF FEMALES"
501 PRINT
510 FOR K=1 TO 16
520 PRINT K,M(K,N),F(K,N)
530 NEXT K
990 END
```

Fig. 6.2 (Cont.)

The number of generations or time periods that the program should run is entered by statement 235. Because of storage limitations, the maximum number of generations was set at 25. On computers with a larger memory this number could easily be raised. By developing the program in a manner similar to that indicated at the beginning of this chapter, the use of double subscripts would have been avoided. This would have greatly reduced the storage requirements and such a procedure may be necessary when using computers with very limited storage capacity. This is an example of how the hardware, or computer capability, may dictate the problem formulation. As computer models become more complex and increase in size, such limitations upon problem formulation become more severe.

The program was run for 25 time periods, and as stated above, assumed an initial population of 100 deer in each age class, both male and female. Figure 6.3a is a print out of the results. These results agree with those of Usher (1972) given in his table 3. The comparison is made by "normalizing" our results at the 25th time period so that the number of males in the first age class is 1000. The normalizing is done by dividing each of the male and female age class populations by 11217.8, and then multiplying the quotients by 1000. The number 11217.8 is the number of males in the first age class at the last time period as is shown by the listing of the population by age class for the last time period in figure 6.3a. The normalized populations are listed in columns 3 and 5 of figure 6.3b and the "unnormalized" populations are listed in columns 2 and 4 of the same figure. The normalized populations have been rounded to the nearest integer.

Since the population growth of the Rhum Island deer population appears to steady down to a constant rate of growth independent of the initial starting population distribution, it is of

```
RBLT1

N=NO. OF YRS. EVOLUTION. N MUST NOT EXCEED 25
?25


                 PROGRAM RESULTS


PERIOD       NEWBORN MALES     NEWBORN FEM.     TOT. POP.

 1               567               612.4             3452
 2               433.416           480.484           3548.04
 3               432.84            484.932           3785.35
 4               513.328           573.444           4232.5
 5               578.659           643.622           4805.91
 6               637.191           770.464           5539.68
 7               727.274           905.307           6458.41
 8               846.39            1062.82           7504.78
 9               996.362           1261.05           8647.92
 10              1157.61           1432.17           9925.71
 11              1340.19           1656.51           11484.6
 12              1565.91           1936.76           13334.9
 13              1821.84           2255.13           15497.3
 14              2118.4            2629.08           18028.1
 15              2466.46           3060.12           20978.4
 16              2871.16           3559.68           24406.7
 17              3339.49           4139.76           28385.6
 18              3885.15           4913.53           33020.7
 19              4520.24           5601.65           38422
 20              5260.03           6520.01           44708.8
 21              6120.3            7586.79           52022.6
 22              7121.17           8827.63           60531.6
 23              8286.03           10270.9           70431
 24              9641.19           11950.2           81948.5
 25              11217.8           13904.3           95350
 26              13052.4           16178.5           110945


POPULATION BY AGE CLASS FOR THE LAST TIME PERIOD
AGE CLASS, NO. OF MALES, NO. OF FEMALES

 1               11217.8           13904.3
 2               6922.37           10313
 3               5889.88           7995.14
 4               5011.25           6060.8
 5               4263.86           4578.6
 6               3627.88           3391.8
 7               3089.58           2447.81
 8               1949.14           1699.56
 9               830.992           741.064
 10              264.348           207.734
 11              192.57            154.294
 12              135.789           109.363
 13              91.2048           75.984
 14              56.4427           47.9637
 15              29.5153           27.8959
 16              9.27994           12.7584
```

Rhum Island Deer Results

Fig. 6.3a

A Comparison of Normalized and Unnormalized Results

N = 25 years

	No. of Males		No. of Females	
K	From Run	Normalized to M(1,N)=1000	From Run	Normalized to M(1,N)=1000
1	11219	1000	13904	1240
2	6922	617	10313	920
3	5890	525	7995	713
4	5011	447	6061	540
5	4264	380	4579	408
6	3628	323	3392	302
7	3090	275	2448	218
8	1949	174	1700	152
9	831	74	741	66
10	264	24	208	19
11	192	17	154	14
12	136	12	109	10
13	91	8	76	7
14	56	5	48	4
15	30	2	28	2
16	9	1	13	1

Fig. 6.3b

interest to examine the growth rate of the populations of individual age groups. We choose to examine the increase or decrease each time period of the number of newborn. This may be done by comparing the ratio of the number of male newborn for two successive time generations at several consecutive generations. For example, the ratio of the number of male newborn in the 25th time period to the number of male newborn in the 24th time period is 1.16. This indicates that the male population has gained 16% from the 24th to the 25th generation. If the corresponding ratio for the female newborn is calculated, the value 1.16 is again obtained. The equality of the two ratios suggests that the population change "has settled down" to a constant change in population for each period and that the magnitude of this change is equal to 16% per time period. To check the assertion that the rates were approaching a limiting value as the number of generations increase, we calculated these same male and female newborn ratios for earlier generations. They are listed in the table below.

N (Generation)	M(1,N)/M(1,N-1)	F(1,N)/F(1,N-1)
5	1.100	1.196
10	1.158	1.157
15	1.1638	1.1634
20	1.1636	1.1637
25	1.1635	1.1636

This table indicates that, indeed the change in population is approaching a constant value of 16% per period. In fact, this value is almost achieved after only ten generations. Further evidence that, as the number of generations increases, the population is settling down or converging to a constant change in population per period may be gained by examining the change in population per period for each age class. Your author did this by choosing the number of time periods, N, in the program to be 24 and 25 respectively and comparing the resultant age class populations. It was noted that the percentage increase for each age class was a constant and equal to 16.36 percent. These results indicate that the change in population appears to have converged to a value approximately equal to 16.36 percent and that this change is the same for each age group. We will denote this value by V.

A precise discussion of convergence is the subject of advanced mathematics. Nevertheless, it is useful to present a short and very heuristic discussion of convergence as it effects our work. In brief, the mathematician would say that by making the number of time periods large enough, it is possible to insure that the change in population from one time period to the next becomes as close as desired to a constant value. Of course, in a computer it is not possible to run the program for an infinite or indefinitely large number of time periods and thus, we must arrive at an approximate result by running for a sufficiently large number of time periods. We can obtain a good estimate for this number by comparing the population change as the number of periods increases. This comparison is done with the finite digit arithmetic of the computer. Consequently, if the change in the population from one generation to the next appears constant for several generations to within the number of digits of the arithmetic used in the computer, we will say that the corresponding value for the population change is the desired value for V.

In an attempt to determine what effect different distributions of starting populations might have on the growth rate, three more runs were made. The first run assumed an initial population distribution of 160 deer in the first age group, 150 deer in the second age group and so on down to the 16th age group which began with only 10 deer. In the second run, the distribution of the starting populations was reversed. Thus, the first age group contained 10 deer, and the 16th age group contained 160 deer. In the last run, the beginning population distribution consisted of 400 deer in the sixth through and including the tenth age gropus and no deer in the other age groups. The program was modified to calculate the quotients P(I+1)/P(I) where the index I counts the time periods. Only these quotients were printed out. Figure 6.3c lists a summary of the results obtained by running the program for 25 time periods. The order of the columns corresponds to the order of the specified initial population distributions as described above. The fourth column lists the same results for the run of the program in which the initial population consisted of 100 deer in each age class.

As noted previously, the population of each age group, as well as the total population, always settled down to a constant rate of growth. Thus, the relative numbers of individuals of different ages remain constant from one time period to the next. A population which has these characteristics is said to be a population which has a stable age distribution.

It is nearly always the case that survivability and fecundity data obtained from populations which have inhabited an area for a long period of time, characterize a stable age group

Growth Rates Corresponding to Different Initial Population Distributions

GR. RATE	GR. RATE	GR. RATE	GR. RATE
1.16348	.994019	1.1195	1.07875
1.0904	.95457	.984213	1.02782
1.10336	1.01812	1.03413	1.06689
1.13121	1.09917	1.13041	1.11813
1.14012	1.12855	1.16912	1.13548
1.15099	1.15523	1.17279	1.15268
1.16168	1.1721	1.16509	1.16584
1.16271	1.16099	1.16447	1.16202
1.15795	1.14394	1.15013	1.15232
1.15419	1.13806	1.14131	1.14776
1.15929	1.15363	1.15466	1.15705
1.16176	1.16012	1.16067	1.16111
1.16246	1.16169	1.1613	1.16216
1.1633	1.16331	1.16254	1.1633
1.16362	1.1637	1.1632	1.16366
1.16352	1.16327	1.1632	1.16342
1.16323	1.16271	1.16267	1.16302
1.16336	1.16317	1.16315	1.16329
1.16355	1.16361	1.16364	1.16357
1.16359	1.16368	1.1637	1.16363
1.16357	1.16361	1.16359	1.16359
1.16356	1.16357	1.16354	1.16356
1.16355	1.16353	1.16352	1.16354
1.16354	1.16352	1.16352	1.16353
1.16354	1.16353	1.16353	1.16353
1.16355	1.16356	1.16356	1.16355

Fig. 6.3c

population. This is because the population has become "adjusted" to its environment which is assumed not to have changed. Most large herd populations have a steady age distribution and this fact is very helpful in obtaining survivability and fecundity data.

By increasing the number of age groups and altering the fecundity and survivability rates, the program can be used to study changes in human populations. Since human populations evolve over a long period of time, it will be necessary to require a much larger computer memory or to redesign the program to minimize the storage requirements. Provision must also be made to alter the rate data during the course of the run. For example, it is quite evident that in many highly industrialized populations, the fecundity rate is decreasing and the mortality rate by age group is slowly changing. The program must have the capability to accept these changing rate data. Because their rate data is not constant over time, such populations do not have stable age structure population distributions.

It is evident that a constant increase in the population each time period will result in an increasingly large population for each age class as the number of time periods increases. The accumulation of large numbers of individuals per age group can be prevented by "normalizing" the age group populations before beginning the life table calculation for the next time period. This is accomplished by setting the population of some reference age group equal to a constant, say 1000, and then "scaling" all of the other populations in accord with this population. For the Rhum Island deer program, it is convenient to choose the number of newborn males as the reference age class. If M1(K,I) and F1(K,I) denote the quotients M(K,I)/M(1,I) and F(K,I)/M(1,I) respectively, then by choosing the values 1000*M1(K,I) and

1000*F1(K,I), K=1 to N, for the initial populations each time period, we have normalized our population. An estimate of the value of V is made by calculating the change in population of the smallest age class each period. The student should carry out the program modifications necessary to effect this normalization.

When the population growth has achieved a stable value, the change in the total population can be described by a single equation:

$$T(I+1) = T(I) + V * T(I).$$

In this equation, T(I) denotes the total population and V denotes the growth coefficient corresponding to a stabilized growth. Thus, V is analogous to G in the Malthus problem and the assumed constancy of the survival and birth rates of the life table model corresponds to the assumed constancy of the birth and mortality rates of the Malthus model. It should be noted that there do exist sets of values for the S1(K) and S2(K), and for the Y1(K) and Y2(K), such that it is not possible to reach a stable age distribution. See Poole (1974). With the aid of equations (7) or (8) in the appendix of the first chapter, it is possible to relate V to the intrinsic growth rate r. This is usually done in mathematically oriented texts on population biology. We will not do so since it is more meaningful to speak of the change in the population over a specified time period, such as a year or a generation, rather than as an instantaneous time rate of change.

The fact that the change in population from one generation to the next approaches a constant value independent of the starting population suggests that the primary quantities of interest in describing the time evolution of a population are the survivability coefficients, S1(K) and S2(K), and the fertility coefficients, Y1(K) and Y2(K). It is frequently assumed that the survival rates of the males and the females are identical and also that the fertility coefficients are equal. Under these assumptions the evolution of the population can be completely characterized by a consideration of only the changes in the female population. Thus, the sets of parameters S2(K) and Y2(K) are sufficient to describe the population growth.

In the literature, these sets of coefficients are frequently referred to as P_x and F_x respectively, and are interpreted in a probabilistic sense, e.g. Poole (1974). Thus, P_x denotes the probability that a female alive at time t, and in the age group from x to $x+1$, will still be alive at time $t+1$ and hence, in the age group from $x+1$ to $x+2$. Here the variable x counts the time periods which may be measured in generations, years, etc. The variable t is the elapsed time and is measured in the same units as is the variable x. The symbol F_x denotes the number of female offspring born at time t, that will still be alive at time $t+1$, of a female in the age group x to $x+1$. Hence, the BASIC variables I and K correspond to the traditional variables t and x respectively.

Let the initial female population of the cohort be denoted by P(0). Since only S2(0) of these will be alive at the end of the first time period, the actual number alive at the end of this period is S2(0)*P(0). Similarly, since S2(1) denotes the proportion surviving the second time period, the number alive at the end of the second time period is S2(0)*S2(1)*P(0). Continuing in this way, it is seen that the number of females from the original cohort that are alive at the end of N time periods is

S2(0)*S(1)*S2(2)* * * *S2(N-1)*P(0).

Thus, the proportion of the original cohort population still alive after N time periods is

S2(0)*S2(1)* * * *S2(N-1).

In the literature, this proportion is usually denoted by l_x. l_x is the proportion of the original population alive at time t and of age x. It is also helpful to introduce another variable, denoted by m_x, which denotes the average number of offspring produced per female whose age is in the interval x to x+1. Hence, the quantity, $l_x m_x P(0)$, is the total number of births from females of age x. If this quantity is summed over all x, that is summed over all age groups, the total number of births is obtained. Consequently, $l_x m_x$ is a measure of the reproductive capacity of the females of age x, and the $l_x m_x$ vs. x curve is a measure of the reproductive capacity of the entire population.

Since the proportion

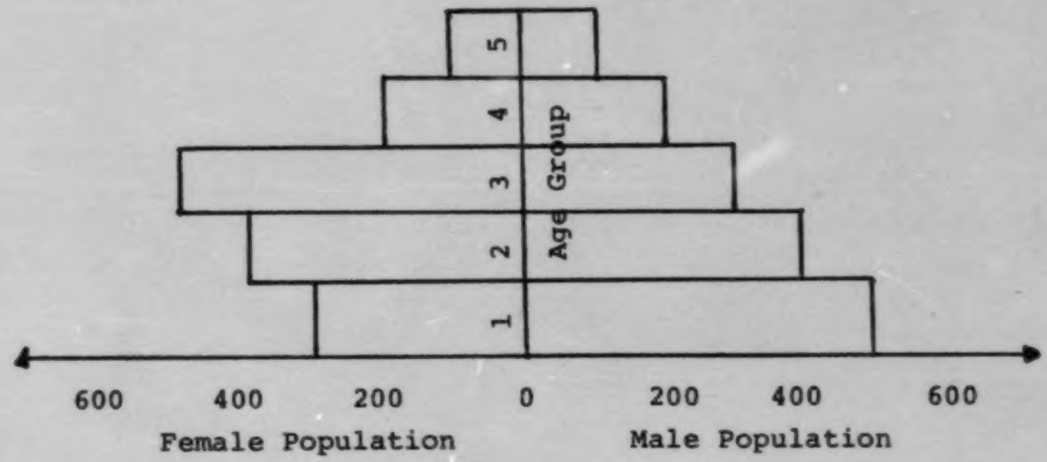
S2(0)*S2(1)* * * *S2(N-1)

is l_x, it is possible to evaluate this quantity for all values of N, and then graph it as a function of N. In this way, the measure of the reproductive capacity of the Rhum Island deer population could be displayed.

The preceding comments were made to enable the student to more easily relate the programming notation to that used in the literature. There is a wide diversity of notation and terminology in the population dynamics literature and so the subject will not be pursued here. Comments on the notation can be found in Mertz (1971).

A Pictorial Representation

A useful technique for displaying the age group population structure at any given time period is a form of horizontal bar graph. The following example will serve to indicate the method of construction of such a graph. Suppose there are only 5 age groups in the population and the male population of each age group is 500, 400, 300, 200 and 100 respectively. Let the female population of each age group be 300, 400, 500, 200 and 100 respectively. The male population distribution is represented by a horizontal histogram on the right of the vertical axis and the female population distribution is represented by a horizontal bar graph on the left side of the vertical axis. In both representations, the base of the bar graph is the vertical axis. Pictorially the population structure appears as shown in figure 6.4.



Population Structure

Figure 6.4

Since the small age groups are on the bottom and the older age groups are on the top, the figure clearly demonstrates that most of the male population is young, whereas most of the female population is middle aged. If the figure were top heavy, the population would consist largely of older members. To enable the ready analysis of the effect of varying survivability, fecundity rates and initial populations, the program shown in figure 6.2 was modified to describe a population of only 5 age groups. Figure 6.5 is a listing of the program.

Line 20 provides for the necessary storage. The survivability rates are entered as input in lines 80 to 120 and lines 115 to 150 provide for the inputting of the fecundity data. The initial population of each age class is entered in lines 175 to 210. By providing the ability to directly enter such data as input, the evolution of populations with very different conditions may be easily examined. The number of generations, N, the program is to be run is entered by lines 230 and 235. The index I counts the generations and the index K counts the age groups. The actual generation by generation calculation begins at line 405 and continues through line 490. Lines 410 and 411 initialize the counters M(1,I) and F(1,I), which count the number of male and female newborn respectively. This initialization is done at the beginning of each generation. The numbers of males and females in the next age group, for the next time period, are calculated in statements 430 and 440, while lines 450 and 460 calculate the number of newborn during the I^{th} time period. Finally, the results are printed out in lines 505 to 590. A print out of a typical run is shown in figure 6.6 and figure 6.7 is a pictorial representation of the growth of the population for the first 5 generations.

```
RBLT5

5 REM           5 AGE GROUP LIFE TABLE PROGRAM
6 REM
7 REM
8 REM
10 PRINT "            5 AGE GROUP LIFE TABLE PROGRAM"
11 PRINT
12 PRINT
13 PRINT
20 DIM M(6,26),F(6,26)
39 REM
40 REM      INPUT MALE SURVIVABILITY RATES
41 REM
45 PRINT "TYPE 5 MALE SURVIVABILITY RATES"
50 INPUT S1(1),S1(2),S1(3),S1(4),S1(5)
55 PRINT
74 REM
75 REM    INPUT FEMALE SURVIVABILITY RATES
76 REM
80 PRINT "TYPE 5 FEMALE SURVIVABILITY RATES"
85 INPUT S2(1),S2(2),S2(3),S2(4),S2(5)
90 PRINT
109 REM
110 REM          INPUT FECUNDITY DATA FOR MALE NEWBORN
111 REM
115 PRINT "TYPE 5 MALE BIRTH RATES"
120 INPUT Y1(1),Y1(2),Y1(3),Y1(4),Y1(5)
125 PRINT
139 REM
140 REM          INPUT FECUNDITY DATA FOR FEMALE NEWBORN
141 REM
145 PRINT "TYPE 5 FEMALE BIRTH RATES"
150 INPUT Y2(1),Y2(2),Y2(3),Y2(4),Y2(5)
155 PRINT
169 REM
170 REM         INPUT INITIAL MALE POPULATIONS
171 REM
175 PRINT "TYPE 5 STARTING MALE POPULATIONS"
180 INPUT M(1,0),M(2,0),M(3,0),M(4,0),M(5,0)
185 PRINT
199 REM
200 REM         INPUT INITIAL FEMALE POPULATIONS
201 REM
205 PRINT "TYPE 5 STARTING FEMALE POPULATIONS"
210 INPUT F(1,0),F(2,0),F(3,0),F(4,0),F(5,0)
215 PRINT
```

Fig. 6.5

```
230 PRINT "TYPE N=NO. OF YRS. EVOLUTION.   N MUST NOT EXCEED 25"
235 INPUT N
240 PRINT
245 PRINT
250 PRINT
255 PRINT "                         PROGRAM RESULTS"
260 PRINT
265 PRINT
395 PRINT "TIME PERIOD, NEWBORN MALES, NEWBORN FEMALES"
396 PRINT
399 REM
400 REM               I COUNTS THE TIME PERIODS
401 REM
405 FOR I=0 TO N
410 M(1, I+1)=0
411 F(1, I+1)=0
414 REM
415 REM               K COUNTS THE AGE CLASSES
416 REM
420 FOR K=1 TO 4
430 M(K+1, I+1)=S1(K)*M(K, I)
440 F(K+1, I+1)=S2(K)*F(K, I)
450 M(1, I+1)=M(1, I+1)+Y1(K)*F(K, I)
460 F(1, I+1)=F(1, I+1)+Y2(K)*F(K, I)
470 NEXT K
480 PRINT I+1, M(1, I+1), F(1, I+1)
490 NEXT I
495 PRINT
500 PRINT
505 FOR I=0 TO N
507 PRINT "THE GENERATION NUMBER IS   "; I
510 PRINT "AGE CLASS, NO. OF MALES, NO. OF FEMALES"
520 PRINT
550 FOR K=1 TO 5
560 PRINT K, M(K, I), F(K, I)
570 NEXT K
580 PRINT
585 PRINT
590 NEXT I
990 END
```

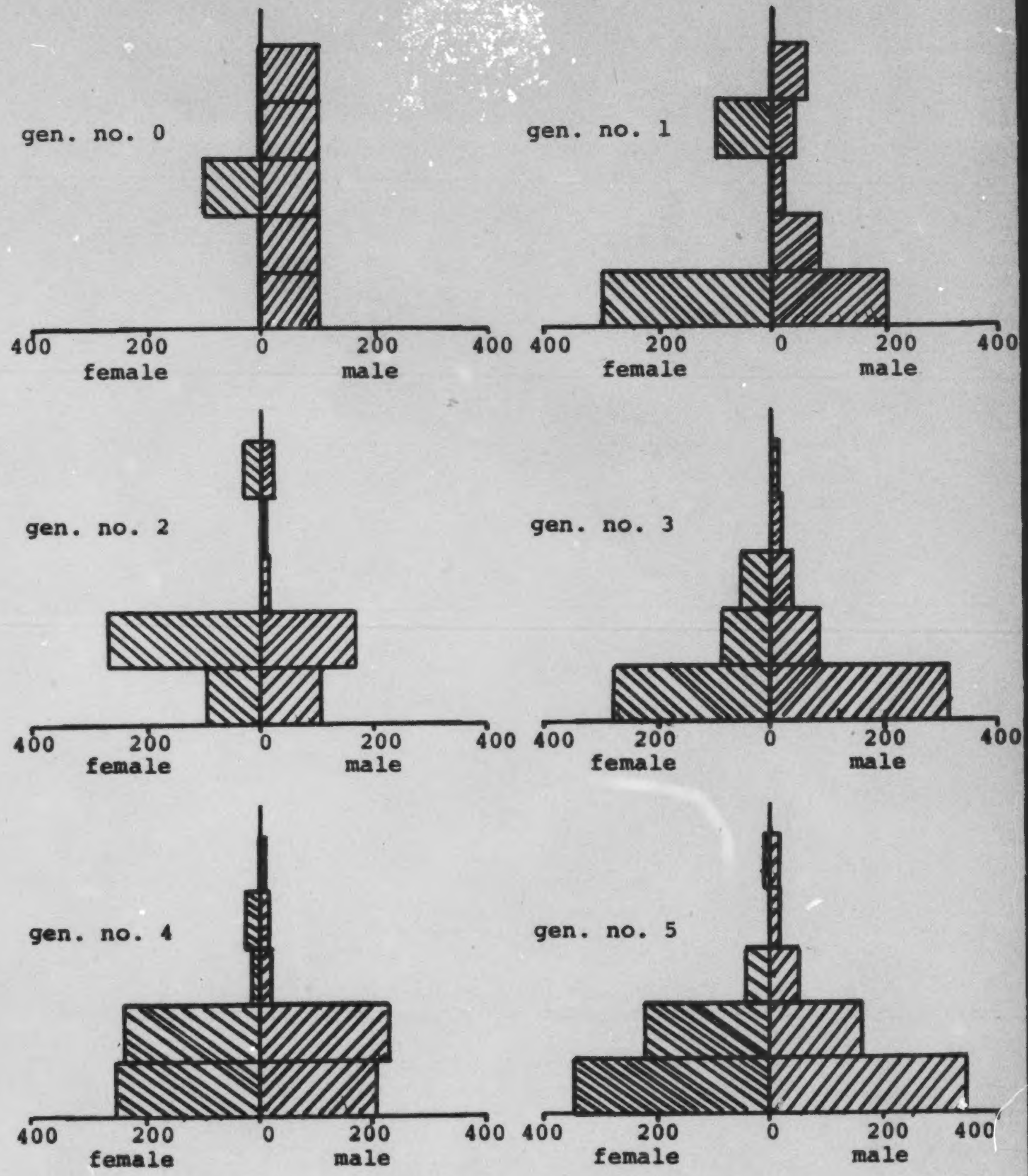
Fig. 6.5 (Cont.)

```
RBLT5

                    5 AGE GROUP LIFE TABLE PROGRAM


TYPE 5 MALE SURVIVABILITY RATES
?.8, .2, .4, .6, 0

TYPE 5 FEMALE SURVIVABILITY RATES
?.9, .2, .5, .6, 0

TYPE 5 MALE BIRTH RATES
?0, 1, 2, 2, 1

TYPE 5 FEMALE BIRTH RATES
?0, 1, 3, 2, 1

TYPE 5 STARTING MALE POPULATIONS
?100, 100, 100, 100, 100

TYPE 5 STARTING FEMALE POPULATIONS
?0, 0, 100, 0, 0

TYPE N=NO. OF YRS. EVOLUTION.   N MUST NOT EXCEED 25
?6


                         PROGRAM RESULTS


TIME PERIOD, NEWBORN MALES, NEWBORN FEMALES

 1             200           300
 2             100           100
 3             270           270
 4             198           252
 5             333           351
 6             342           390.6
 7             455.22        500.58


THE GENERATION NUMBER IS    0
AGE CLASS, NO. OF MALES, NO. OF FEMALES

 1             100           0
 2             100           0
 3             100           100
 4             100           0
 5             100           0
```

Fig. 6.6

```
THE GENERATION NUMBER IS     1
AGE CLASS, NO. OF MALES, NO. OF FEMALES

 1             200          300
 2             80           0
 3             20           0
 4             40           50
 5             60           0

THE GENERATION NUMBER IS     2
AGE CLASS, NO. OF MALES, NO. OF FEMALES

 1             100          100
 2             160          270
 3             16           0
 4             8            0
 5             24           30

THE GENERATION NUMBER IS     3
AGE CLASS, NO. OF MALES, NO. OF FEMALES

 1             270          270
 2             80           90
 3             32           54
 4             6. 4         0
 5             4. 8         0

THE GENERATION NUMBER IS     4
AGE CLASS, NO. OF MALES, NO. OF FEMALES

 1             198          252
 2             216          243
 3             16           18
 4             12. 8        27
 5             3. 84        0

THE GENERATION NUMBER IS     5
AGE CLASS, NO. OF MALES, NO. OF FEMALES

 1             333          351
 2             158. 4       226. 8
 3             43. 2        48. 6
 4             6. 4         9
 5             7. 68        16. 2

THE GENERATION NUMBER IS     6
AGE CLASS, NO. OF MALES, NO. OF FEMALES

 1             342          390. 6
 2             266. 4       315. 9
 3             31. 68       45. 36
 4             17. 28       24. 3
 5             3. 84        5. 4
```

Fig. 6.6
(Cont.)



Evolution of Population Structure

Figure 6.7

Harvest

The previous discussion concerning the Rhum Island deer herd population suggests that it should be possible to harvest a proportion of the population and yet permit the population to remain constant or even increase. There are several ways to harvest a population, and the simplest method, which we will call the equal harvest method, is to harvest the same proportion, H, from each age class. The inclusion of an equal harvest in our program is accomplished by noting that during each generation the number of males and females harvested from each age class is `H*M(K,I)` and `H*F(K,I)`. This number of males and females must be removed from our population each time period and so the age group equations become

```
430 LET M(K+1,I+1)=S1(K)*M(K,I)-H*M(K,I)
```

and

```
440 LET F(K+1,I+1)=S2(K)*F(K,I)-H*F(K,I).
```

The calculation of the number of newborn is not changed since it is assumed that they are not harvested. Provision must be made for "inputting" the harvest proportion, H.

The allowable range of values for H is restricted by the minimum value of the male and female survivability coefficients. The harvest proportion must be less than this minimum value to insure the survivability of the succeeding age classes. H must also be greater than or equal to zero since H=0 corresponds to no harvest. We also assume that survivability and fecundity rates are not affected by harvesting. Figure 6.8 illustrates results obtained from the harvest program. The results were obtained for harvest rates of 0.1, 0.163, 0.2 and 0.5 respectively. In each case the program was run for ten years and the initial population was that listed in figure 6.3a, columns 3 and 5. By choosing the initial population in this way, we are assuming that the population growth of the herd has "settled down" to a constant rate of change. Thus, we are mimicing the harvesting of a herd that is accustomed and adjusted to its habitat. By comparison with the initial population, it is seen that a harvest rate of 0.1 permitted an increase in the population for all age classes whereas a harvest rate of 0.163 resulted in almost no change in the age class population. The harvest rates of 0.2 and 0.5 respectively resulted in moderate and severe decreases in the population. All of the final numerical results appearing in the tables have been rounded to the nearest integer because integer arithmetic was not used in the program. By using different harvest ratios for each age class, several harvest policies can be explored.

It is of interest to discuss the implications of harvesting with a harvest proportion equal to V, the limiting value of the population change. We assume that harvest takes place after the emergence of the newborn and that no newborn are harvested. Under these conditions, the population will remain stable, that is, there will be no change in the total population from one period to the next. The student should verify this assertion by altering the life table program to include harvesting, setting H=V, and using as initial population the normalized age group populations corresponding to V. Since both V and the normalized populations have been determined numerically,

HARVEST PROGRAM RESULTS

AGE GROUP	H=0.1		H=0.163		H=0.2		H=0.5	
	MALE	FEMALE	MALE	FEMALE	MALE	FEMALE	MALE	FEMALE
1	1855	2292	997	1229	678	832	11	13
2	1084	1656	560	872	370	583	4	8
3	912	1256	467	649	307	430	3	5
4	768	928	391	472	256	309	3	3
5	647	683	327	342	213	221	2	2
6	546	492	275	240	178	153	2	1
7	460	345	230	165	149	104	1	1
8	274	231	131	108	83	66	0	0
9	103	90	44	38	26	22	0	0
10	27	20	10	7	5	3	0	0
11	18	13	6	4	3	2	0	0
12	13	9	4	3	2	1	0	0
13	8	6	3	2	1	1	0	0
14	5	4	2	1	1	0	0	0
15	2	2	1	1	0	0	0	0
16	1	1	0	0	0	0	0	0

POPULATION BY AGE CLASS AFTER HARVESTING

WITH DIFFERENT RATES FOR 10 YEARS.

THE INITIAL POPULATION IS GIVEN IN COLUMNS 3 AND 5 (FIG. 6.3a)

FIG. 6.8

and because of the subsequent finite digit arithmetic, there will be a slight drift in the results. However, the population will remain relatively unchanged for several time periods. By varying H and noting the change in the population distribution by age class over several time periods, it is possible to estimate the value of H which produces a prescribed, but constant, change in the population. If the value of H is such that no change is produced over several generations, then H=V. Such a scheme can actually serve as an effective iteration scheme for the determination of V.

V is related in a simple manner to the value of the latent value, λ , mentioned by Usher in his article. The relation is

$$\lambda \doteq 1 + V$$

where $\doteq$ denotes approximate equality. If N_K denotes the population in the K^{th} age class corresponding to V, then the exact relation between λ and V is

$$\lambda = 1 + V - VN_M/(\sum_{K=1}^{M} N_K) \qquad (6.1)$$

Here M denotes the maximum number of age classes, and male and female age classes are considered as distinct. The term

$$\sum_{K=1}^{M} N_K$$

denotes the sum of the normalized age group populations. The correction term

$$VN_M/(\sum_{K=1}^{M} N_K)$$

is quite small. Using the Rhum Island deer herd data, taking $M=32$, and $V=0.1636$, the value of the correction term is found to be $1.9x1.0^{-5}$. Thus, the relation $\lambda = 1+V$ then is quite accurate. If the newborn had also been harvested in the same proportion, the relation $\lambda = 1+V$ would have been exact.

We purposely did not derive this result since the derivation uses linear algebra. The result is mentioned to indicate the connection between our elementary approach and the more mathematically sophisticated approach of Leslie and Usher. The student who is familiar with linear algebra will find it instructive to derive the result. The effectiveness of the matrix formulation of the problem is due to the fact that the elements in the matrix are assumed to be constant. Since many of the useful results of matrix algebra rely on the fact that the elements of the matrix are constant, this formulation is somewhat limited in its applicability. In contrast, the structure of the BASIC language age group equations permits them to include survival and fecundity rates which can vary from one time period to the next. Such alterations can be easily effected in the program. Other typical alterations might include the effect of finite resources, the effect of contamination, prey-predator interactions, the effect of competition, the effect of time lags, etc.

Discussion

Integer arithmetic was not used in the calculation because your author felt that the consequent increase in computer time was not worth the possible benefit. On the other hand, the use of integer arithmetic when outputting data would have resulted in a more realistic appearing population and probably should have been included in the program.

An examination of the table of newborn males and females shows that as the number of time periods increases, the number of newborn continues to increase. This suggests that the deer population would become arbitrarily large after a sufficient number of time periods. Since this is an impossibility, it must be the case that there exists a factor in their environment that results in survivability and fecundity rates which imply an increasing population. It is probably the case that this factor is harvesting, i.e. hunting pressure. Thus, assuming equal harvest, we determined that approximately 16% of the herd was harvested each year. Your author has not checked the validity of this assertion with the actual hunting harvest. Nevertheless, he suspects that this is the case. It is certainly the case that if no plausible reason can be discovered for the continued growth of the herd then the experimental data is in error. The purpose of the above discussion was to illustrate another way in which theory can assist experiment.

In order to make the development easy to follow, the computer programs developed in this section did not contain any safety checks. If the programs were to be actually used, they should contain safety checks. For example, it is possible that the population of an age class may drop below a single individual. In this event the population for that age class has effectively vanished and provision should be made for setting it equal to zero. If population effects are included (see next section), and the modification parameters are chosen too large, the populations of an age class may even become negative. Provision must be made in the program to test for such an occurrence. The existence of negative populations or abnormal population changes are signs of possible errors in the data or in the program. For these and similar reasons, it is imperative that the programmer analyst give reasonable effort to ascertaining the possible existence of physically unrealizable results. He then must make provision in the program for testing for their occurrence, and in the event such results do occur, the program must have the proper instructions to enable it to take the correct action.

Population Effects (Effect of a Finite Resource)

It is known that fecundity rates as well as survivability rates are decreased by an increasing population living in a finite environment. In this section, we discuss the inclusion of such an effect which is sometimes called the population effect. As an aid in this discussion, it is helpful to note the analogy of the life table development with the previous development of the exponential growth model. In each model, the growth rates were assumed to be constant and independent of population. Now, it should be recalled that the necessary modification of the exponential growth model to include the effect of a finite environment was accomplished by modifying the average growth rate by an amount proportional to the present population. This suggests modifying the life table development in a similar manner.

We will assume that the survivability rates, as well as the fecundity rates, for each age class, are reduced by an amount proportional to the total population. Thus, if $P(I)$ denotes the total population at the beginning of the I^{th} time period, the survivability rates must be written as

$$S1(K)-S3(K)*P(I) \qquad \text{and} \qquad S2(K)-S4(K)*P(I).$$

The sets of values $S3(K)$ and $S4(K)$, $K=1,2,,,,,N$ are constants of proportionality relating the reduction in the survivability rates to the present population. They are called the survivability modification parameters and are positive and very much smaller than each of the $S1(K)$ and $S2(K)$.

The accounting for the effect of the finite environment on the fecundity rates is accomplished in a like manner. Thus, the birth rates are written as

$$Y1(K)-Y3(K)*P(I) \qquad \text{and} \qquad Y2(K)-Y3(K)*P(I)$$

where Y3(K) and Y4(K), K=1,2, ,,, ,N are two sets of proportionality constants relating the decrease in the fecundity rates to the present population. Y3(K) and Y4(K) will be called the fecundity rate modification parameters. The total population, P(I) is given by

```
10 LET P(I) = 0
20 LET K=0 TO M
30 LET P(K) = P(K)+M(K,I)+F(K,I)
40 NEXT K
50 LET P(I) = P(M)
```

where M is the number of age classes.

The programming alterations necessary to accommodate these changes are very straightforward. Provision must be made for storing the constants, equations 430 to 460 must be changed by the insertion of the modified survivability and fecundity rates, and finally provision must be made for calculating the total population each time period.

Your author made these corrections to the program shown in figure 6.2 and made some sample runs assuming a common value for each of the sets S3(K), S4(K), Y3(K) and Y4(K). The variation of the number of newborn males was used as a measure of the effect of varying the survivability modification parameters. The results are presented in figure 6.8 where it is noted that, as the magnitude of the modification parameters increases, the number of newborn males decreases. It should also be noted that the modification parameters result in a leveling off of the number of newborn. Hence, these parameters are analogous to the growth rate modification parameter G1 that appears in the finite resource model. In these runs, the fecundity rate modification

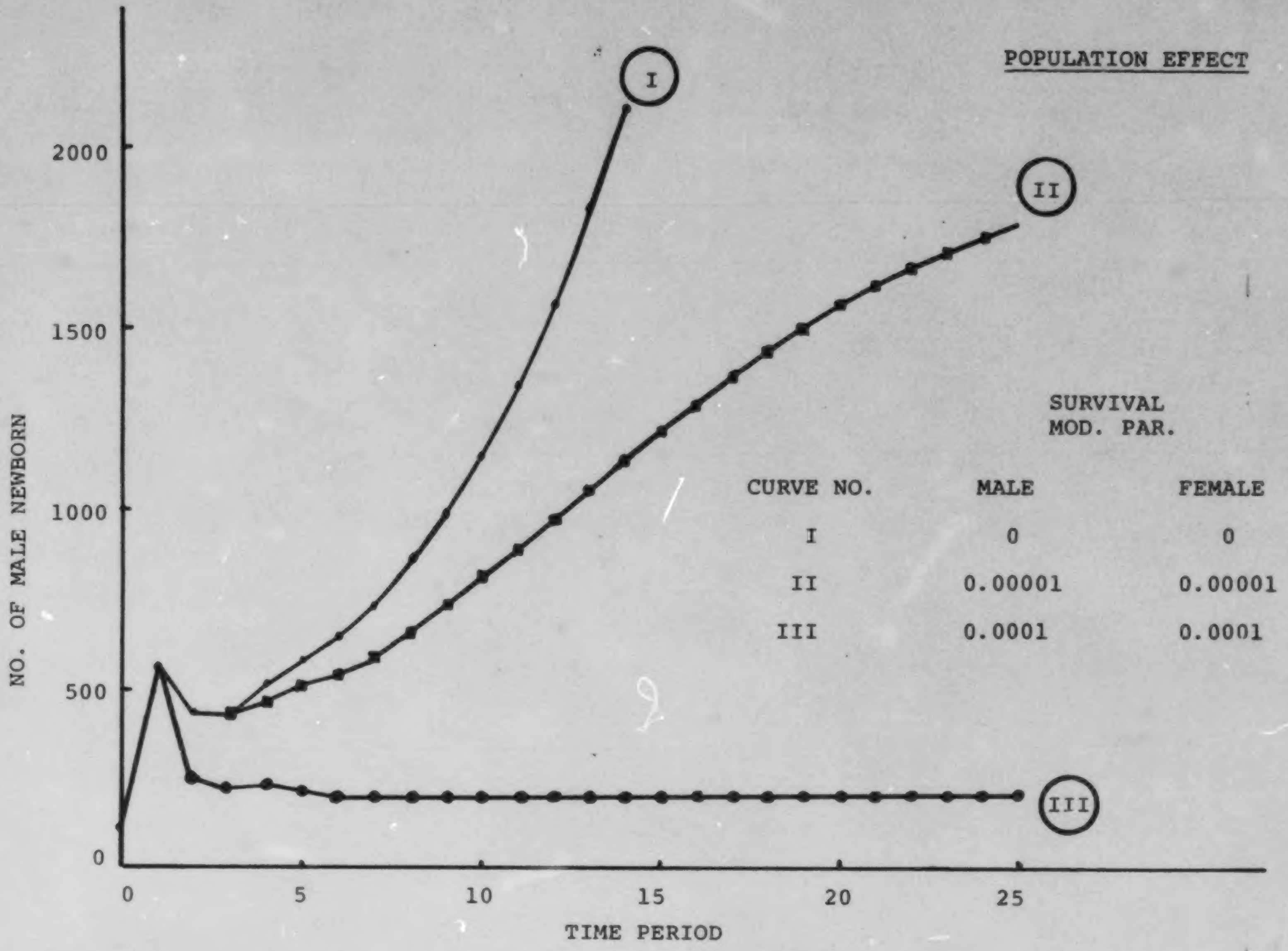

POPULATION EFFECT
I
II
III
NO. OF MALE NEWBORN
2000
1500
1000
500
0
0
5
10
15
20
25
TIME PERIOD
SURVIVAL
MOD. PAR.
CURVE NO. MALE FEMALE
I 0 0
II 0.00001 0.00001
III 0.0001 0.0001

Fig. 6.9

parameters are set equal to zero. If the survivability rate modification parameters are set equal to zero and the fecundity rate modification parameters are varied in an analogous manner, a similar behavior is observed. Many other variations of the parameters may be studied, and other measures of the change in the population, such as the change in the total population, can be used to measure the effects of these variations. The student is encouraged to so modify the program and to then carry out a computer assisted experimental analysis of the effects of parameter variations.

It is important to note that the actual obtaining of the 4N parameters is a nearly impossible experimental task. Thus, the alteration of the program to provide for these constants may not appear fruitful. Nevertheless, a careful examination of results obtained by varying these parameters can result in a better understanding. The previous development again illustrates the fact that the ability to construct very general programs is both a blessing and a curse. (See page 2.22). It is a curse because it tends to negate the real effort and ingenuity required to obtain valid and useful experimental data and, it tends to encourage unjustifiable curve fitting and parameter juggling. On the other hand, it is a blessing because greater insight can frequently be obtained and sometimes such flexibility, when properly combined with empirical knowledge, can permit the determination of the necessary parameters.

Effect of Mating Possibility

The original development of the life table assumed that the ability of a female to produce an offspring was independent of the number of males in the population. Promiscuity is a characteristic of most populations and hence, this is a reasonable assumption. On the other hand, there do exist a few populations for which mating partners, once chosen, do remain faithful throughout the life span of each partner. For such monogamous populations it is known that an unequal distribution of the population of each sex can directly affect the number of newborn. Examples of such populations are Geese and Swan. It is the purpose of this section to present a modification of the life table analysis to include the effect of unequal numbers of males and females. In this connection, it may be helpful for the student to reread the section entitled, "Effect of Mating Possibility" in the first chapter. The development of the model will be based upon an attempt to mimic the growth of a population that begins with individuals only in the first three age classes. Thus, initially the first age class consists of the newborn, the second age class is made up of yearlings, the third age class contains two-year-olds who are possible parents and there are no individuals in the remaining (N-3) age classes.

It will be assumed that the time interval for each age class is one year and that the population breeds only once a year. Thus, it is convenient to choose the time period for the life table as one year. It will also be assumed that no individual is older than N years. It is further assumed that the fecundity and survival rates by sex are known for each age class and that finite environment or population effects are negligible. Because an individual is assumed to be monogamous throughout its lifetime, the number of possible mothers in an age class is given by the minimum of the number of males or females in an age class. Finally, it is assumed that the individuals constituting a mating pair select each other the first year of mating possibility. Thus, the fact that the population is monogamous implies that no older individuals will mate with younger individuals, and hence, the individuals in each pair will always remain the same age.

As a result of these assumptions, the number of possible mothers in an age class is given by the minimum of the number of males and females in the age class. The implementation of these assumptions is accomplished by adding the lines

```
442 IF F(K,I)>M(K,I) GOTO 450
444 LET F(K,I)=M(K,I)
```

to the program listed in figure 6.2. It was assumed that there were 100 males and females in each of the first three age classes while the remaining age classes contained no animals.

Using the Rhum Island red deer survivability and fecundity data, your author so modified the program given in figure 6.2. The program was run for 20 periods and the results are given in figure 6.10 in columns 3 and 5. The original Rhum Island red deer program listed in figure 6.2 was modified to have the same initial age class population distribution and also run for 20 time periods. The results of this program are listed in columns 2 and 4 of figure 6.10. It is seen that the modification due to mating possibility resulted in a decrease in the number of newborn as would be expected. Similar conclusions can be inferred with respect to other comparisons of population distributions. The inclusion of population effects may be accomplished by adjoining to the program those statements developed in the previous section.

MATING POSSIBILITY

TIME PERIOD N	NEWBORN MALES		NEWBORN FEMALES	
	ORIG. PROG.	MOD. PROG.	ORIG. PROG.	MOD. PROG.
0	100	100	100	100
1	62	62	66	66
2	94	91	99	96
3	107	101	129	123
4	118	111	153	146
5	138	124	182	168
6	167	142	215	189
7	195	156	241	199
8	225	171	275	216
9	262	191	320	242
10	305	218	378	279
11	355	248	442	319
12	413	279	513	360
13	481	313	597	401
14	560	350	693	447
15	651	392	806	500
16	757	441	938	563
17	881	496	1092	634
18	1025	558	1271	715
19	1193	627	1479	803
20	1388	704	1720	901

Figure 6.10

PROBLEMS

CHAPTER VI

1. Alter the initial age class population and run the program for 24 generations or so.
2. Alter survivability and/or fertility rates, then make runs to see how sensitive results are to such changes.
3. Rewrite the program on figure 6.2 using integer arithmetic. Make some trial runs and compare the results with those obtained from the original program.
4. Figure 6.11 lists life table data on the White River elk herd in Colorado and is presented with the courtesy of Dr. R. Ream of the University of Montana School of Forestry. The herd is heavily harvested and has a very high reproductive rate. For purposes of herd management, it is convenient to designate the following classes of animals:
    - (1) Calves - including both males and females less than one year old.
    - (2) Cows - includes all females in the population except calves.
    - (3) Spike bulls - includes all yearling males.
    - (4) 2-5 point bulls - includes all males in 2, 3, and 4 year old age class.
    - (5) 6-point bulls - includes all males 5 years and older. These are called the trophy bulls.

    (a) Modify the program given in figure 6.2 to simulate the time evolution of the age group populations of the elk herd.

    (b) It is assumed that the harvestors (hunters) can distinguish between the five classes. It is further

ELK POPULATION TABLE

Based on White River Herd
Data -- Colorado

AGE	NO. OF MALES	MALE SURVIVAL	NO. OF FEMALES	FEMALE SURVIVAL	REPRODUCTIVE RATE
0	1000	.82	1000	.82	0
1	800	.95	800	.95	0
2	240	.90	700	.95	.95
3	65	.85	550	.90	.98
4	35	.80	320	.85	.96
5	25	.70	200	.75	.95
6	9	.65	120	.70	.90
7	3	.65	66	.70	.90
8	1	.60	40	.65	.85
9	1	.50	22	.60	.80
10	0	.40	12	.60	.75
11	0	.30	7	.55	.70
12	0	.20	4	.50	.60
13	0	.20	2	.50	.50
14	0	0	1	.50	.40
15	0	0	0	.40	.30
16	0	0	0	.30	.20
17	0	0	0	0	0

Figure 6.11

assumed that each hunter is successful and will harvest a single animal in that class for which he is given a license. How should licenses be issued to:

(1) Maximize the number of trophy bulls?

(2) Maximize the total number of animals taken?

5. Figure 6.12 lists survivability and population data for the Yellowstone Park Grizzly Bear population as obtained from Craighead, et al, 1973. The fecundity rate was found to be approximately 1.00 for the years 1959-67 and 0.68 for the period 1968-70. These rates are constant over all but the first five age classes and these classes cannot give birth. Construct a life table model of the bear population making provision for acceptance as input the sex ratio of the newborn and also the fecundity rate. Make some runs with different birth rates and initial populations. Both the survivability and the fertility rates are in error due to sampling limitations. The sensitivity of the population to these errors can be found by making computer runs with different values for these rates.

6. As suggested in the section describing population effects, modify the Rhum Island deer program to include the effect of increasing population. State the assumptions behind your modifications. Make up some data and run the program. Discuss your results.

7. Using the program listed in figure 6.5, make up a set of runs to examine the effect on the evolution of the population due to:

   (a) Changes in the male survivability rates only,

   (b) Changes in the female survivability rates only,

   (c) Changes in both,

   (d) Changes in the initial population, both male and female, and

   (e) Any combination of changes you select.

Age	Number in Age Class	Number Males	Survivorship Rate Px	Number Females	Survivorship Rate Px
0.5	33.0	19.5	0.7436	13.5	0.6296
1.5	23.0	14.5	0.6828	8.5	0.9529
2.5	18.0	9.9	0.8586	8.1	0.6790
3.5	14.0	8.5	0.8235	5.5	0.9091
4.5	12.0	7.0	0.5143	5.0	0.8200
5.5	7.7	3.6	0.9444	4.1	0.9756
6.5	7.4	3.4	0.9412	4.0	0.9500
7.5	7.0	3.2	0.9688	3.8	0.9737
8.5	6.8	3.1	0.9677	3.7	0.9730
9.5	6.6	3.0	0.9667	3.6	0.9444
10.5	6.3	2.9	0.9655	3.4	0.9706
11.5	6.1	2.8	0.9643	3.3	0.9394
12.5	5.8	2.7	0.8889	3.1	0.9032
13.5	5.2	2.4	0.8750	2.8	0.8571
14.5	4.5	2.1	0.7619	2.4	0.7917
15.5	3.5	1.6	0.7500	1.9	0.7368
16.5	2.6	1.2	0.8333	1.4	0.8571
17.5	2.2	1.0	0.8000	1.2	0.7500
18.5	1.7	0.8	0.7500	0.9	0.8889
19.5	1.4	0.6	0.8333	0.8	0.7500
20.5	1.1	0.5	0.8000	0.6	0.6667
21.5	0.8	0.4	0.7500	0.4	0.7500
22.5	0.6	0.3	0.6667	0.3	0.6667
23.5	0.4	0.2	0.5000	0.2	0.5000
24.5	0.2	0.1	0.5000	0.1	0.5000
25.5	0.1	0.1	0.0000	0.1	0.0000
TOTALS	178.0	95.4		82.7	

Grizzly Bear Data

Fig. 6.12

State your variations and the reasons for choosing them. Discuss and compare the results. Do they agree with your intuition?

8. Make a run with the 5 group program using rate data and starting populations of your own selection. Run the program for N1 generations. Record the populations and then use these populations as initial populations for a new run, of N2 generations, in which the rate data are changed. Examine your results. Discuss them. Note that this procedure of "chaining" runs corresponds to changing the rate data during the course of one long run.

9. Alter the 5 group program to accept changing rate data. Make up your own data and carry out some runs. Discuss your results.

REFERENCES

CHAPTER VI

Craighead, J. J., Varney, J. R. and Craighead, F. C., Jr., 1973. A Computer Analysis of the Yellowstone Grizzly Bear Population. Montana Cooperative Wildlife Research Unit, University of Montana.

Emlen, J. M., 1973. Ecology: An Evolutionary Approach. Addison-Wesley. Reading, Massachusetts.

Keyfitz, N., 1968. Introduction to the Mathematics of Population. Addison-Wesley. Reading, Massachusetts.

Leslie, P. H., 1945. On the Use of Matrices in Certain Population Mathematics. Biometricka 34, 183-212.

Mertz, D. B., 1971. Life History Phenomena in Increasing and Decreasing Populations, in Statistical Ecology, Volume 2, Sampling and Modeling Biological Populations and Population Dynamics. Ed. by Patil, G. P., Pielou, E. C. and Waters, W. E. The Pennsylvania State University Press, University Park, Penn.

Poole, R. W., 1974. An Introduction to Quantitative Ecology. McGraw-Hill. New York, NY.

Usher, M. B., 1972. Developments in the Leslie Matrix Model. In Mathematical Models in Ecology. pp. 29-60. Ed. by J. N. R. Jeffers. Oxford.

CHAPTER VII

APPLICATIONS TO GENETICS

Some Preliminaries

The science of genetics is that body of knowledge which attempts to explain the transmission of physiological and mental characteristics from the parents to their offspring. Biologists believe that the unit of structure of a living organism is the cell and that it is through the cell that the characteristics of an organism are expressed. This expression is accomplished by collections of rod shaped bodies called chromosomes which themselves are made up of sets of smaller bodies called genes. It is these latter bodies which biologists believe determine the characteristics or hereditary traits transmitted from the parents to the offspring. It is known that a certain number of genes from each of the parents unite to form a unique collection of genes in the offspring and since there is associated with each gene or set of genes a characteristic or characteristics, an accounting of the distribution of the genes in the offspring will enable us to specify the hereditary traits or characteristics of the offspring. Consequently, in this chapter we will attempt to use a computer to determine and/or list the number and kinds of genes in the cell of the offspring. We will assume that the student is somewhat familiar with the biology of heredity; at least to the extent of having had a secondary school, elementary college or university level course in the life sciences.

Thus, we will freely use some elementary biological terms early in our discussion. We will begin by describing the work of Gregor Mendel who in 1865 founded the modern science of genetics. Most students are familiar with his work and know that he postulated or stated laws that:

(a) Permitted the assigning of definite probabilities to specified gene distributions that occurred in the cells of the offspring, and

(b) associated the gene distribution of the offspring with a distribution of heredity traits and thus enabled the prediction of the heredity characteristics of the offspring.

Since some students may not be familiar with Mendel's work it will be described in the next section.

Mendel's Experiments

Our discussion will begin with a summary of Mendel's experiments. The summary will include only that part of his work with which the student must be familiar in order that he understand the work that is to follow. In the event that our presentation of Mendel's work is too brief, the student is urged to consult an elementary modern biology text or some of the references at the end of the chapter.

Mendel worked with the pea plant. His experiments consisted of crossing and growing successive generations of the plants. For each successive generation, he would note the biological or physical characteristics of the parent plants, cross plants with specific characteristics with other plants of the same or different set of characteristics, and then tally the distribution of these characteristics in the offspring. For example, if the characteristic of interest was height of the plant, Mendel would cross several pairs of plants, each member of which was a tall plant, each member was a short plant, or the members of the pairs would be such that one was short and the other tall. By carefully listing the distribution of the characteristic of the offspring and then comparing this distribution with the specified distribution of the same characteristic in the parents, Mendel was able to devise a hypotheses which enabled him to predict, with great accuracy, the distribution of the characteristic in the offspring resulting from a specified distribution of this characteristic in the parent generation.

He worked with several pure bred varieties of pea plants. Each variety of plant was distinguished from the other by such outwardly distinguishable characteristics as vine height, position of flowers, color of seed, color of flowers, etc. In total, Mendel noted the pea plants possessed seven distinct biological characteristics which carried over from generation to generation. Furthermore, each of these characteristics was characterized by the fact that it could be in only one of the two possible states. For example, the characteristic of color of seeds could either be in the yellow state (a yellow seed) or in the green state (a green seed), and the characteristic of stem height could either be in the tall or the short state. Mendel recognized that in order to establish the validity of his hypothesis concerning the distribution of the states of the characteristics through succeeding generations, that he would have to have a collection of seeds with the property that when they were pollinated (regenerated) solely among themselves they would always reproduce in their offspring the unique characteristic states of the parents. Plants with this capability are said to "breed to type" or are pure bred. Mendel also recognized that it would be mandatory to be able to absolutely control the pollination or fertilization of each plant in order to rigorously account for the matings. This was readily accomplished in the pea plant by the well-known technique of artificial fertilization or pollinization.

Because the plants were pure bred, the self-pollinization of two distinct tall plants would produce a tall plant. Analogous results would occur after the breeding of two pure bred small plants. However, the breeding by artificial pollinization of a tall plant with a short plant (such breeding is called hybrid breeding and the offspring labeled hybrids), also produced a tall plant regardless of which plant the original pollen came from. Since the state of tallness always resulted from such a mating, Mendel had established experimentally the fact that the state of tallness was dominant as compared to the state of shortness. He thus labeled tallness a dominant state or trait and shortness a recessive state or trait. It is important that the student recognize that Mendel had established this relation only for the pea plant. The very opposite could well be true for some other plant.

Having established the relation of dominance for the state of tallness over the state of shortness, Mendel then experimented with crosses involving other characteristics of the pea plant. In a very similar manner, he found by experiment that by crossing plants that produced yellow seeds with plants that produced green seeds that all first generation plants had yellow seeds and thus yellow was the dominant seed color as compared to green. Similarly, he found that round seeds were dominant to wrinkled seeds. After performing analogous experiments to compare all seven characteristics, Mendel found the surprising result that for each of the seven characteristics one of the states of each characteristic was always dominant. Stated in another way, one of the states of each characteristic appeared to be lost.

The following table is a summary of Mendel's findings:

The Pairs of Contrasting States	Dominant State of the F_1 Hybrid Offspring
Rounded seeds, wrinkled seeds	Round seeds
Yellow seeds, green seeds	Yellow seeds
Colored seed coat, white seed coat	Colored seed coat
Inflated pod (unripe), constricted pod	Inflated pod
Green pod, yellow pod	Green pod
Axial flowers, terminal flowers	Axial flowers
Tall stem, short stem	Tall stem

Having discovered the existence of dominant traits or states by breeding pure bred types, Mendel then proceeded to experimentally determine the distribution of these characteristic states in succeeding generations. As a means of keeping track of each generation, he labeled the hybrid offspring of the two different but pure bred plants the first filial or $\underline{F_1}$ generation, the offspring of these offspring the second filial or $\underline{F_2}$ generation, and so on for each succeeding generation.

Because both parents of the original parents were pure bred, the offspring of either self-pollinated tall or self-pollinated short plants were indeed either all tall or all short, respectively. However, it was the resultant distribution of the height characteristic states occurring in the offspring of the F_1 generation whose parents were hybrids for the states of tallness and shortness that was most unexpected and historically significant. Mendel found that approximately 75% of the plants were tall and 25% were short even though all of the parents were tall. Furthermore, upon experimenting in a similar way with the other characteristics he found analogous distributions for the states of these characteristics. Thus, other F_2 generations consisted of plants in which nearly 75% were green colored and very nearly 25% were yellow colored; in still another set of F_2 generation plants about three-fourths had round seeds and about one-fourth had wrinkled seeds. These results were obtained despite the fact that the parents were all green colored or all had round seeds. The fact that the proportions were so definite and consistent was indeed most surprising and Mendel set himself the task of devising a hypothesis for the rational explanation of such behavior. It was this hypothesis and consequent explanation which earned him the title of "The Father of Modern Genetics".

His reasoning was simple and yet imaginative and is one of the best examples in science of a mental model or hypothesis to explain a scientific phenomena. Mendel hypothesized in the following manner in order to establish a model with which to explain the height distribution of the offspring of the hybrid F_1 plants. He assumed the existence, in the parent plant, of a pair of unknown influences which controlled the dissemination of the height characteristic to the offspring. He called these unknown influences, *factors*; however, today we know them as *genes* and will so designate them in the discussion which follows. We also now know that the process of dissemination to the offspring of the parental characteristics is accomplished by the actual giving to the sperm nuclei or *gamete* a single gene from each parent. Furthermore, we know that a pair of gametes, one from each parent, unite to form a single cell called a *zygote*. The zygote is the original cell, which by division and duplication, develops into the offspring plant or organism. The modern theory of heredity holds that there exists in the membrane of the nucleus of the cell a number of distinct linear threadlike bodies, called chromosomes, which are the carriers of the hereditary factors or genes. The division of the cell is called cytokinesis and the events and mechanisms involved in the division of the nucleus of the cell are called mitosis (pronounced my-toe-sis). These events and mechanisms assure that the newly formed cells receive the same number and kind of chromosomes, and hence, the same distribution of genes, as existed in the parent cell.

Now, since some of the hybrid offspring plants in Mendel's experiments were either tall or short, he further assumed that these unknown genes must occur in pairs. In making the assumption of the existence of such pairs of genes for each state, Mendel established his first law of heredity which states:

> "The various hereditary characteristics (such as height, color of plant, shape of seeds, etc.) are controlled by genes and furthermore these genes occur in pairs."

He formulated his second law of heredity by assuming that the tall plants of the F_1 hybrid offspring were unlike the tall plants of the offspring from pure bred tall plants since the state of smallness did not appear in the first or F_1 generation, but would reappear in the next generation. Thus, his second law of heredity states:

> "One gene in a pair of genes may mask or prevent expression of the other gene."

Mendel further assumed that when the gene pairs of the offspring are created that these gene pairs contain only one gene from each of the pair of corresponding genes of each parent. Thus, his third law of heredity, frequently known as the Law of Segregation states:

> "Only one member of any pair of genes in a parent is transmitted to each offspring."

By examining successive generations of offspring of pea plants that originally differed from each other in two characteristics, Mendel established the *Law of Independent Assortment* which states:

> "A gene pair associated with one characteristic (for example, height) segregates independently of a gene pair, associated with another characteristic (for example, color of seeds)."

With these laws Mendel was able to predict the distribution of the traits in successive generations of offspring. This prediction was accomplished by calculating the possible pairings of genes of the parents. In order to make such calculations more orderly and more readily understood by the student, it is convenient to introduce some definitions and notations. These are used frequently in describing genetic phenomena and many students are probably familiar with them. As stated previously, we are assuming that for each characteristic such as height, color of seed, etc., there exists in *each* of the parents a heredity factor or set of heredity factors, which specifies a state of the characteristic. Such a heredity factor is called an *allele*. Thus, the gene for tallness and the gene for shortness are alleles of each other. Biologists frequently say that "alternative forms of the same gene are called alleles of each other". Also, the term *allelic forms* is used to describe the existence of the two (or more) heredity factors associated with a given characteristic. If the allele of tallness is dominant as compared to the allele of shortness, the tall allele is said to be dominant over the short allele, or is the *dominant* allele. A *recessive* allele is defined in a corresponding manner. It is customary to label the dominant allele with a capital letter or letters and the recessive allele with a small letter or letters. Thus, the allele corresponding to tallness will be labeled T and the allele for shortness labeled t. When both members of the pair of alleles are the same (for example, TT or tt), the cell is said to be *homozygote.*

Conversely, when the alleles differ, the cell is said to be *heterzygote*. We have seen that some F_1 offspring of a hybrid pairing may all have the same outward appearance of tallness but when these F_1 offspring are paired, they produce some offspring which are short in outward appearance and thus some of the F_1 offspring must have carried a recessive allele of shortness. Consequently, the hereditary constitution of the F_1 offspring was different than their external appearance and so it is convenient to distinguish between these two properties. Thus, the appearance type of the organism will be called the *phenotype* of the organism and the gene or heredity structure of the organism will be called the genotype of the organism.

The assumption of associating a specific gene with a specific biological or physical characteristic is very limited. In fact, it is usually the case that the state of the characteristic, as well as the characteristic itself, is specified by more than the alleles from just a single gene. It is now believed that the principle means by which hereditary characteristics are determined are the chromosomes and their expression in terms of DNA molecules. Genes are assumed to be allігned in a linear order along the chromosome and the location on the chromosome of a specific gene is called the *locus*. It has also been determined that the biochemical processes required in the formation of the basic cells directly influence the development of the chromosomes. The field of study relating these processes to an understanding of heredity is called *cytogenetics*. Developments in this discipline have shown that the gene may not be the ultimate unit of organization. However, for the purposes of this work, it will be assumed that the gene is the fundamental unit of heredity. The hypothesis of the existence of genes and their role in the determination of hereditary phenomena is a very valuable and fruitful hypothesis. This work shall be limited to an elementary analysis of this role.

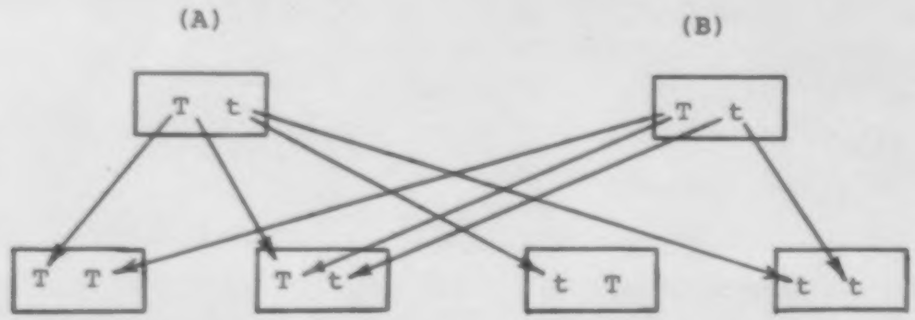
The Prediction of Heredity Characteristics

In this section we will consider the problem of predicting the distribution of the height characteristic of the pea plant offspring knowing the gene distribution of the parent plant. For purposes of illustration, it is assumed that each parent is a hybrid for the characteristic of height. Thus, the chromosome of each parent contains an allele for tallness, as well as an allele for shortness. Mendel's Law of Segregation does not specify which allele of the pair of parent alleles is transmitted by a parent to the gamete of the offspring. Consequently, it is assumed that either allele of each parent may be transmitted and that there exists no internal, nor external, factor prohibiting, nor discriminating in favor of, the transmission of a particular allele of the pair. (Here we are excluding externally induced mutations and the effects of selectivity). Thus, either allele may be transmitted with equal probability and this fact is a fundamental assumption in the following discussion.

We are now in a position to describe the gene dissemination process by a flow diagram. The symbol

$$\boxed{Tt}$$

will denote the gene corresponding to the characteristic of height and the letters T and t denote the allelic forms corresponding to the states of tallness and shortness respectively. The flow diagram or inheritance chart of two hybrid for height parent plants is:



A box in the top row designates the allelic form of the height gene in a parent whereas a box in the bottom row designates one possible pairing, in an offspring, of an allele from each parent. One such pairing results from the transferring of a tall allele from each parent to give a TT allelic pair in the offspring. Another offspring pairing results from the transferring of a T allele from parent A in conjunction with a transferring of a t allele from parent B. The same allelic offspring pairing is obtained by the transferring of a t allele from parent A together with a T allele from parent B. Finally, a fourth allelic pairing in the offspring could arise from the passing of a short allele from both parents. Each of the four possible offspring pairing occur with equal probability because it is assumed that it is equally likely that a tall or a short allele will be passed down from either parent plant. The fact that each of the four possible pairings is equally probable means that for a large number, say 1000, of matings there will be approximately 250 of each pairing. Since the T allele is dominant and it is assumed there is no sexual distinction of alleles, i.e. tT and Tt give equivalent phenotypes, there will be about 750 tall offspring and about 250 short offspring. Thus, there will be very nearly three times as many tall plants as there are short plants. In this way, Mendel explained the 3:1 distribution of tall and short plants, the 3:1 distribution of green and yellow colored plants, etc.

Such pictorial representations provide a basis for the calculation of the prediction of occurrence of a specific gene distribution. In this work, we will not develop calculation methods in a manner that is normally followed in a genetics course; rather we will use a computer to simulate the process (that is we will grow our own plants on the computer) and count the number of desired genotypes. In brief, our simulation will consist of using the random number generator to select "at random" an allele from each parent and to then record the selected allelic pair. (The possible allelic pairs are TT, Tt or tT and tt). After the process has been repeated several times (corresponding to the mating of several pairs of plants) the recorded frequency of occurrence of each allelic pair will permit the determination of the ratio of occurrence of each gentoype and hence the probability of occurrence of tall plants as compared to short plants. Geneticists call the ratio of occurrence, of each genotype to the total number of occurrences of all genotypes, the *genotypic* ratio and the ratio of occurrence of a phenotype to the total number of occurrences of all phenotypes the *phenotypic* ratio.

In the preceding example, the phenotypic ratio for tall plants was three-fourths and for short plants it was one-fourth. The corresponding genotypic ratios for TT, Tt and tt respectively were one-fourth, two-fourths (or one-half), and one-fourth.

The more skeptical student may well ask, "Isn't the process of repeated use of the random number generator and the counting of allelic pairs a rather wasteful procedure when we can get the answer by paper and pencil together with the application of known formulae from statistics and probability?" Our answer is a qualified "Yes". For many simple problems this is certainly the case. However, in attempting to describe the change in gene distribution, which occurs over several generations, when such changes are modified by genetic selection or mutation due to environmental factors, the simple but computationally feasible technique of simulating the "passing of the genes" with the aid of a computer may well be the *only* technique that enables us to describe and predict the resultant hereditary characteristics. There are many, many other situations in which the determination of the apparent final distribution of genes is only possible by modeling and computer-assisted counting. Frequently, both computer modeling and mathematical statistics are used together to carry out the determination. This confrontation between mathematical statistics and a computer is again referred to below. Hence, as in the case of our computer modeling of population phenomena, we shall find that we can, with the aid of suitable models and the computer, analyze far more complex hereditary phenomena than we could if we were limited solely to the tools of mathematical statistics. All we are really doing is taking advantage of the tremendous capacity of the modern electronic digital computer to perform very detailed, repetitive calculations in incredibly short periods of time. The prospective serious life scientist should not lose sight of the premise that his goal is the understanding, predicting and explaining of phenomena in the life sciences in an as unambiguous a manner as is possible, be it with the aid of the language called BASIC, FORTRAN, etc. or with the aid of the language called mathematics or with the aid of both languages. A second and possibly more important reason for proceeding in the above manner is that the very act of modeling or simulating the genetic flow via a computer forces the student to understand thoroughly what is happening in a biological sense. Also, there is a great deal of feedback from the development of the computer simulation to the understanding of the student. In contrast, frequently the mathematical modeling of the phenomena leads to mathematical equations or problems each of which requires a "mathematical trick" to effect the solution. Consequently, a great deal of the student's effort is devoted to discovering these "mathematical tricks" (i.e. taking courses in mathematics) when this effort could be devoted to the further understanding of genetics.

Lastly, from the point of view of probability theory, there is much to be said for this procedure of growing the offspring on a computer. The student who is familiar with the subject of probability and statistics will recall that the statement, "the probability of a four occurring on the single toss of a die is one-sixth" can be interpreted to mean that if the die is tossed a sufficiently large number of times, that the ratio of the number of outcomes which are the occurrence of a four to the total number of outcomes, can be made arbitrarily close to one-sixth. In a similar manner, by repeating the simulation, a sufficiently large number of times, we should expect (in the sense of probability theory) to be able to determine the genotypic ratios.

The approach to the analysis of random processes is a direct carrying out of the results of the "frequency interpretation" definition of the probability of an event.

Of course, it is to be expected, and in fact, it is well documented, that a judicious combination and use of the tools of mathematical statistics in conjunction with the computer does indeed enable us to attack even more complicated genetic phenomena. Thus, though we are deliberately neglecting the use of the mathematical tools of probability theory and statistics for pedagogical reasons, the prospective serious student of genetics is urged to become knowledgeable in these disciplines. However, as stated in the preface, the goal of this text is to introduce the student to another tool, the computer, as an aid in his understanding of his subject, and thus we are deliberately maximizing the assistance of the computer and minimizing the use of formal mathematics.

265

The Development of the Computer Program

We turn now to the development of a computer program which will calculate the number of genotypes arising from N matings in a pure hybrid population. In order that the student may more easily correlate the discussion with the BASIC programming language, which does not have upper and lower case letters, we shall introduce a slight change in notation wherein we shall denote the alleles of a gene by A and B. Also, the development will be restricted to a consideration of only two alleles since the extension to multiple alleles will be apparent. Finally, the discussion will also include only one characteristic and later the extension to more than one characteristic will be indicated.

To begin the discussion of the development of a computer program to simulate the process of transferring genes from the parents to the offspring, we consider the simple, but useful, problem of simulating the genotypical population resulting from the mating within a pure hybrid population. Thus, we wish to simulate the occurrence of a large number of crossings of a male AB parent with a female AB parent assuming equal likelihood of an A or B allele being transmitted from either parent. The labeling male and female is introduced for convenience in distinguishing among parents. Because the parent population is assumed to be pure hybrid, all males and females in the parent population are the AB genotype. Thus, there are no AA nor BB genotypes in either the male or female parent group. For ease of understanding, it is convenient to imagine that there are an equal number of males and females.

The simulation is based upon an idea which is very simple. A random number generator will be used to choose an allele from each parent and the allelic pair will be examined to determine the resulting genotype. The process will be repeated a "large" number of times in order to simulate the genotypic population arising from a large number of such hybrid matings. The number of AA, AB and BB "offspring" genotypes will be recorded and the ratio of each of these three numbers to the total number of matings will be called the genotypic ratios, and will be a description of the genotypical population.

The student will recall that the subroutine which generates or produces a random number produces a positive number whose magnitude is less than or equal to unity. Thus, if R denotes the random number so generated, $0<R<1$. Furthermore, it is equally probable that any number between 0 and 1 may be chosen. Hence, in order to choose an allele at random we can let all random numbers greater than one-half correspond to A alleles and all random numbers less than or equal to one-half be B alleles. We furthermore note that the different genotypes may be distinguished if we introduce two new variables M and F, and let M and F have the values 0 and 1 respectively depending upon whether a B or an A allele was selected from each parent. Thus, if $M+F=2$, then both M and F were one and an A allele was selected from each parent. The resultant offspring genotype was then an AA. Similarly, $M+F=1$ denotes an AB or BA genotype and $M+F=0$ signifies a BB genotype. (An AB genotype is the same as a BA genotype since we are assuming no sexual distinction between alleles). Now, our procedure for determining the genotype of one offspring will be to use the random number generator and to then decide whether the male allele is A or B. The process will be repeated again using the random number generator to decide whether or not the female allele is A or B. In each case, the values of M and F are determined and the quantity $M+F$ is then tested to see whether it has the value 2, 1 or 0, i.e. whether the genotype of the offspring is AA, AB or BB.

The entire calculation is portrayed in the flowcharts, figures 1a and 1b. Figure 1a is a flowchart utilizing a verbal description to portray the order and flow of the program. Figure 1b is a flowchart portraying the same information but utilizing the programming language BASIC to describe the flow. The numbers

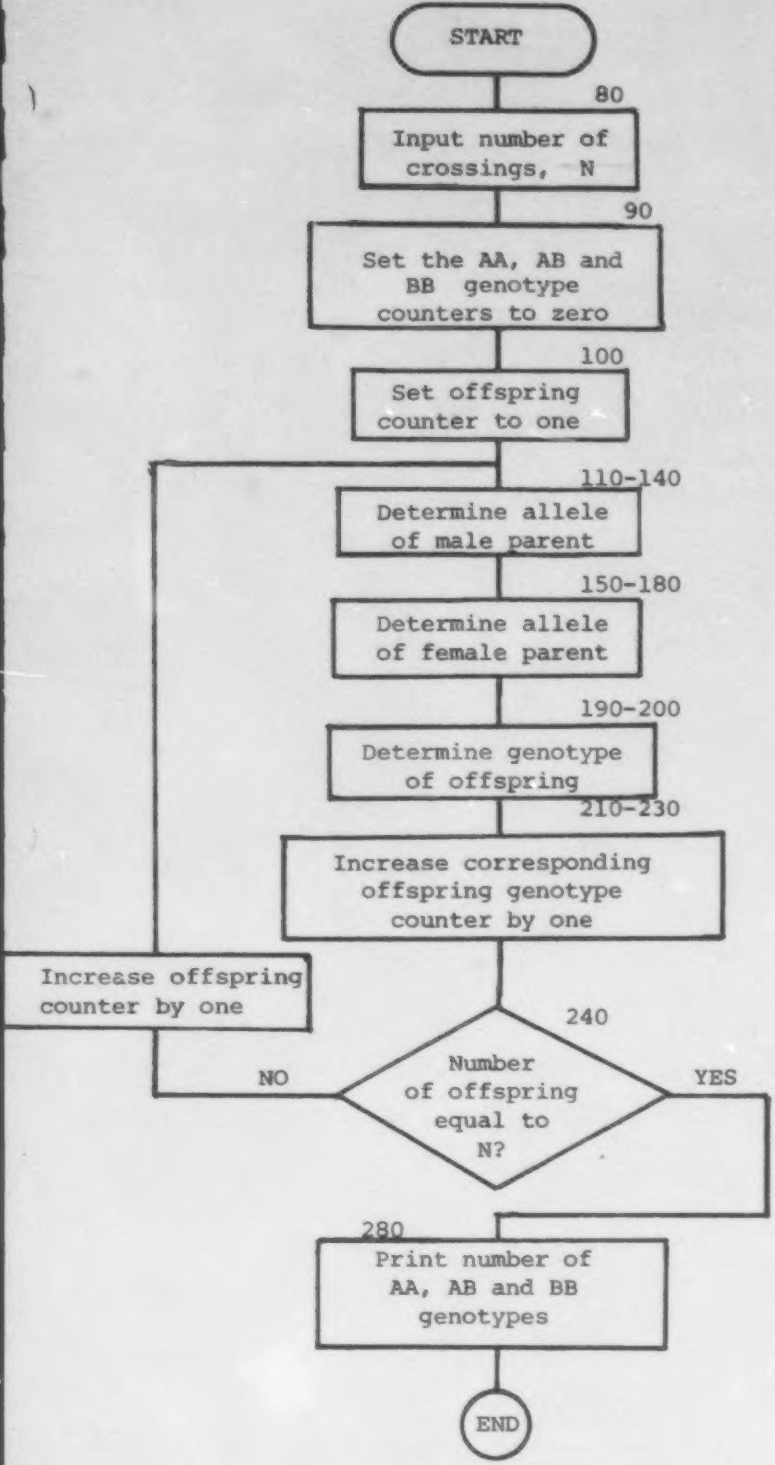

START
80
Input number of crossings, N
90
Set the AA, AB and BB genotype counters to zero
100
Set offspring counter to one
110-140
Determine allele of male parent
150-180
Determine allele of female parent
190-200
Determine genotype of offspring
210-230
Increase corresponding offspring genotype counter by one
Increase offspring counter by one
240
Number of offspring equal to N?
NO
YES
280
Print number of AA, AB and BB genotypes
END

Figure 1a

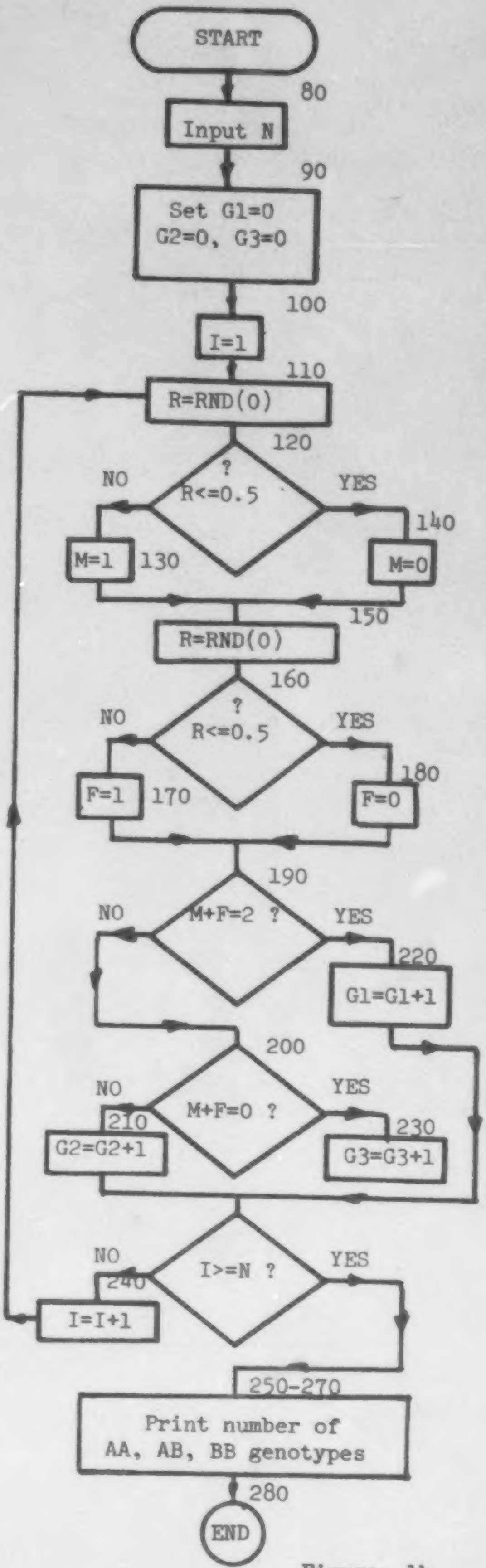

START
80
Input N
90
Set G1=0 G2=0, G3=0
100
I=1
110
R=RND(0)
120
? R<=0.5
NO
YES
M=1
130
140
M=0
150
R=RND(0)
160
? R<=0.5
NO
YES
F=1
170
180
F=0
190
M+F=2 ?
NO
YES
220
G1=G1+1
200
M+F=0 ?
NO
YES
210
G2=G2+1
230
G3=G3+1
I>=N ?
NO
YES
240
I=I+1
250-270
Print number of AA, AB, BB genotypes
280
END

Figure 1b

appearing beside the enclosures refer to the corresponding line of the program which is shown in figure 2. In the program the following notation has been introduced:

G1, G2 and G3 denote the respective number of AA, AB and BB offspring genotypes.

N is the number of matings and it is assumed that each mating always results in just one offspring.

The student should run this program several times and each time increase the magnitude of N the number of offspring. As N gets larger, the student should compare the ratios of the number of AB and BB genotypes to the number of AA genotypes. What do you notice? These ratios are also called the genotypical ratios.

```
GENE

5 REM         FIRST GENETICS PROGRAM
6 REM
10 REM     PURE HYBRID MATING,     SINGLE GENERATION
11 REM
30 RANDOMIZE
50 PRINT "TYPE THE NUMBER OF OFFSPRING"
55 INPUT N
56 PRINT
57 PRINT
64 REM
65 REM   LINES 70 TO 90 SET THE GENOTYPE COUNTERS TO ZERO
66 REM
70 LET G1=0
80 LET G2=0
90 LET G3=0
91 REM
92 REM      LINES 110 TO 180 DETERMINE MALE AND FEMALE ALLELE
93 REM
95 REM      LINES 190 TO 230 DETERINE AND COUNT GENOTYPES
96 REM
100 FOR I=1 TO N
110 LET R=RND
120 IF R<=.5GO TO 140
130 LET M=1
135 GO TO 150
140 LET M=0
150 LET R=RND
160 IF R<=.5GO TO 180
170 LET F=1
175 GO TO 190
180 LET F=0
190 IF M+F=2GO TO 220
200 IF M+F=0GO TO 230
210 LET G2=G2+1
215 GO TO 240
220 LET G1=G1+1
225 GO TO 240
230 LET G3=G3+1
240 NEXT I
250 PRINT "THE NUMBER OF AA GENOTYPES IS";G1
260 PRINT "THE NUMBER OF AB GENOTYPES IS";G2
270 PRINT "THE NUMBER OF BB GENOTYPES IS";G3
275 PRINT
280 PRINT "THE PROPORTION OF AA GENOTYPES IS";G1/N
290 PRINT "THE PROPORTION OF AB GENOTYPES IS";G2/N
300 PRINT "THE PROPORTION OF BB GENOTYPES IS";G3/N
400 END
```

Fig. 2

Discussion of Some Results

Figure 3 illustrates the results of several runs for which the initial population was varied from 25 to 10,000. An examination of these results suggests that, as the number of crossings is increased, the genotypic ratio approaches 0.25 : 0.50 : 0.25, that is 1:2:1. This is the genotypic ratio that is assumed by the offspring population resulting from a purely random mating of an infinite number of hybrid matings. This result can be derived by a simple counting argument.

Because the process is random, one should not expect that, if these runs were duplicated, the exact same results would be obtained. There would be a variation from these results. In order to get a feel for the possible degree of variability of the results, three sets of three runs each were made with the same number of crossings in each run of a set. The number of crossings was chosen to be 50, 100 and 1000 respectively. The results are summarized in Table 1. Considerable variation is noted in the set of runs which correspond to 50 crossings and a lesser degree of variation is displayed by the results in which 1000 crosses were used. This shows that the degree of variation of the genotypic ratios decreases as the number of crossings increases. This is to be expected since small samples usually possess a greater degree of variability than large samples.

The student should be aware that, if he attempts to run this program on the computer in order to verify the correctness of his version of the program, identical results will not be obtained. This is due to the fact that different brand computers have different random number generator subroutines and also have varying degrees of arithmetic accuracy. In addition, line 30 RANDOMIZE insures (at least to within the finite capacity of the computer) that a different sequence of random numbers will be generated each time the program is run. For some versions of the BASIC programming language, the omission of the RANDOMIZE

```
RUN

GENE

TYPE THE NUMBER OF OFFSPRING
?25


THE NUMBER OF AA GENOTYPES IS 10
THE NUMBER OF AB GENOTYPES IS 12
THE NUMBER OF BB GENOTYPES IS 3

THE PROPORTION OF AA GENOTYPES IS .4
THE PROPORTION OF AB GENOTYPES IS .48
THE PROPORTION OF BB GENOTYPES IS .12

READY


RUN

GENE

TYPE THE NUMBER OF OFFSPRING
?50


THE NUMBER OF AA GENOTYPES IS 11
THE NUMBER OF AB GENOTYPES IS 26
THE NUMBER OF BB GENOTYPES IS 13

THE PROPORTION OF AA GENOTYPES IS .22
THE PROPORTION OF AB GENOTYPES IS .52
THE PROPORTION OF BB GENOTYPES IS .26

READY


RUN

GENE

TYPE THE NUMBER OF OFFSPRING
?100


THE NUMBER OF AA GENOTYPES IS 30
THE NUMBER OF AB GENOTYPES IS 50
THE NUMBER OF BB GENOTYPES IS 20

THE PROPORTION OF AA GENOTYPES IS .3
THE PROPORTION OF AB GENOTYPES IS .5
THE PROPORTION OF BB GENOTYPES IS .2

READY
```

```
RUN

GENE

TYPE THE NUMBER OF OFFSPRING
?500


THE NUMBER OF AA GENOTYPES IS 116
THE NUMBER OF AB GENOTYPES IS 249
THE NUMBER OF BB GENOTYPES IS 135

THE PROPORTION OF AA GENOTYPES IS .232
THE PROPORTION OF AB GENOTYPES IS .498
THE PROPORTION OF BB GENOTYPES IS .27

READY


RUN

GENE

TYPE THE NUMBER OF OFFSPRING
?1000


THE NUMBER OF AA GENOTYPES IS 255
THE NUMBER OF AB GENOTYPES IS 506
THE NUMBER OF BB GENOTYPES IS 239

THE PROPORTION OF AA GENOTYPES IS .255
THE PROPORTION OF AB GENOTYPES IS .506
THE PROPORTION OF BB GENOTYPES IS .239

READY


RUN

GENE

TYPE THE NUMBER OF OFFSPRING
?10000


THE NUMBER OF AA GENOTYPES IS 2499
THE NUMBER OF AB GENOTYPES IS 5063
THE NUMBER OF BB GENOTYPES IS 2438

THE PROPORTION OF AA GENOTYPES IS .2499
THE PROPORTION OF AB GENOTYPES IS .5063
THE PROPORTION OF BB GENOTYPES IS .2438

READY
```

Fig. 3

GENOTYPIC VARIATION

		Number of Crossings								
		50			100			1000		
No. of Genotypes	AA	.26	.22	.30	.29	.22	.25	.243	.264	.268
	AB	.58	.58	.52	.51	.49	.53	.489	.496	.479
	BB	.16	.20	.18	.20	.29	.22	.268	.240	.233

Table 1

statement insures that the same sequence of random numbers will be generated each time the program is run. Such a capability is of great assistance when debugging a program which makes use of a random number generator. This is due to the fact that the detection of a presumed program bug in a program is greatly facilitated if the same sequence of operations can be assured each time the program is run. This assurance is obtained if the same set of random number is used each time the program is run.

The success of this program in providing insight concerning the distribution of genotypic ratios resulting from pure random mating in an infinite hybrid population is dependent upon the generation of a truly random sequence of numbers. Such a sequence is necessary because the random mating of an infinite population is being simulated by a finite number of supposedly random matings. Heuristically speaking, a sequence of numbers is said to be random if the occurrence of each number is equally likely and if there is no pattern in the order in which the numbers appear. A more precise definition of a random sequence of numbers requires ideas from advanced statistics and need not be discussed here.

The generation of sequences of random numbers by a computer requires an algorithm for doing so. Since the algorithm can be repeated, identical sequences of numbers can be generated. In contrast, if a toss of a die were used to generate a sequence of random numbers, the repetition of the toss of the die to generate a second sequence, would not result in the generation of the same sequence of random numbers. Consequently, the numbers generated by a random number generator are usually called *psuedo-random numbers*. Nevertheless, the algorithms used in such subroutines produce sequences of numbers which satisfy, or nearly satisfy, many of the sophisticated tests for randomness.

It is also evident that the limited word length of the computer restricts the number, N, of distinct random numbers that can be generated. Thus, the range of the numbers appearing in such sequences are restricted. Because of this, the user must be careful about the use of the same random number generator for generating sequences containing very very many numbers. If the use of such sequences is contemplated, inquiry should first be made of the computer center personnel to ascertain the advisability or feasibility of doing so. These comments are heuristic, and for the most part, constitute part of the lore surrounding the use of Monte-Carlo methods. Monte-Carlo is the name given to techniques or methods which require a sequence, or sequences, of random numbers. Such methods may also be called stochastic methods.

Simple Extensions

The previous program is quite limited in scope and after studying the program in conjunction with the flowchart, the student will easily see how to modify it in order to make it more general and hence useful. To illustrate to the student how readily the program may be generalized, we give some extensions. Certainly it seems unreasonable to restrict ourselves to the simulation of genotypes arising from matings in only a pure hybrid population. Thus, we will alter the program so that the investigator, or user of the program, may specify the number of AA, AB and BB genotypes in the original population. In the altered program these numbers will be denoted by A1, A2, and A3 respectively. Since random mating will be assumed, it will be necessary to simulate the random selection of pairs from the genotypic ratios implied by the relative population of each parent genotype. For ease of presentation, it will be further assumed that the genotypic ratios of both the male and the female parents are the same. This means that the probability of an AA, AB or a BB male or female parent being selected as one of the partners in the mating is proportional to the respective number of AA, AB or BB parents.

The student should note that by properly choosing the number of AA, AB and BB genotypes in the original parental population, it is possible to specify the genotypic ratios of the original parent population. The specification of the genotypic ratios by specifying the respective genotypic populations is possible because the population is finite. Such a method of specification would not be possible if the population were infinite in size. In that event, the genotypic ratios would have to be directly specified. However, because of the finite capacity and ability of the computer, an infinite population cannot be simulated and the specification of the genotypic ratios in terms of the respective numbers of genotypes is possible. An original genotypic ratio of 2:5:3 can be specified in an infinite number of ways. For example, an original parental genotype population of 200AA, 500AB and 300BB, or a population of 100AA, 250AB and 150BB or an original population consisting of 2AA, 5AB and 3BB would each imply an original genotypical ratio of 2:5:3.

Another extension of the program is suggested by the fact that it is of interest to be able to examine the genotypic ratios over several generations. This may be accomplished by assuming that the offspring are the sole parents of the succeeding generation. Thus, the respective numbers of offspring genotypes in a given generation will be used to determine the genotypic ratios of the parents of the subsequent generation. It will also be assumed that the genotypical distribution of the offspring, when acting as parents for the next generation, is the same for both sexes. In another program modification (to be described later) it will be seen that this restriction is easily removed. It will again be assumed that if a parent is an AB genotype that is equally likely, that an A or a B allele will be transmitted. (See for reference instruction numbers 380-403 and 440-463 on page 7.30).

It is important the student understand that the process of not altering the genotypic ratios of the parents as each offspring is created requires that the parent population be infinitely large. In a finite population, the selection of even a single pair of parents does indeed alter the genotypic ratios of the remaining set of parents. However, if the parental population is very large, as we are assuming, the change in the genotypic ratios is so small that it can be ignored. Such a change cannot be ignored when the population is small. Thus, in the program modification under discussion it is assumed that the population is so large that the mimicing of matings may be accomplished without the necessity of accounting for the alterations of the parental genotypic ratios after each mating. An analogous assumption is made in probability theory (see the chapter entitled, Random Processes) when the simulation of random events with the aid of a random number generator is used to calculate probability estimates when the ratio of the sample size to the population size is very small. It is more correct to think in terms of a parent gene pool consisting of three genotypes. Our discussion has been couched in terms of parents and their associated genotypes in order to present the ideas as simply as possible. Genetics is a very sophisticated subject and to even attempt a complete quantitative discussion in this work would be out of place.

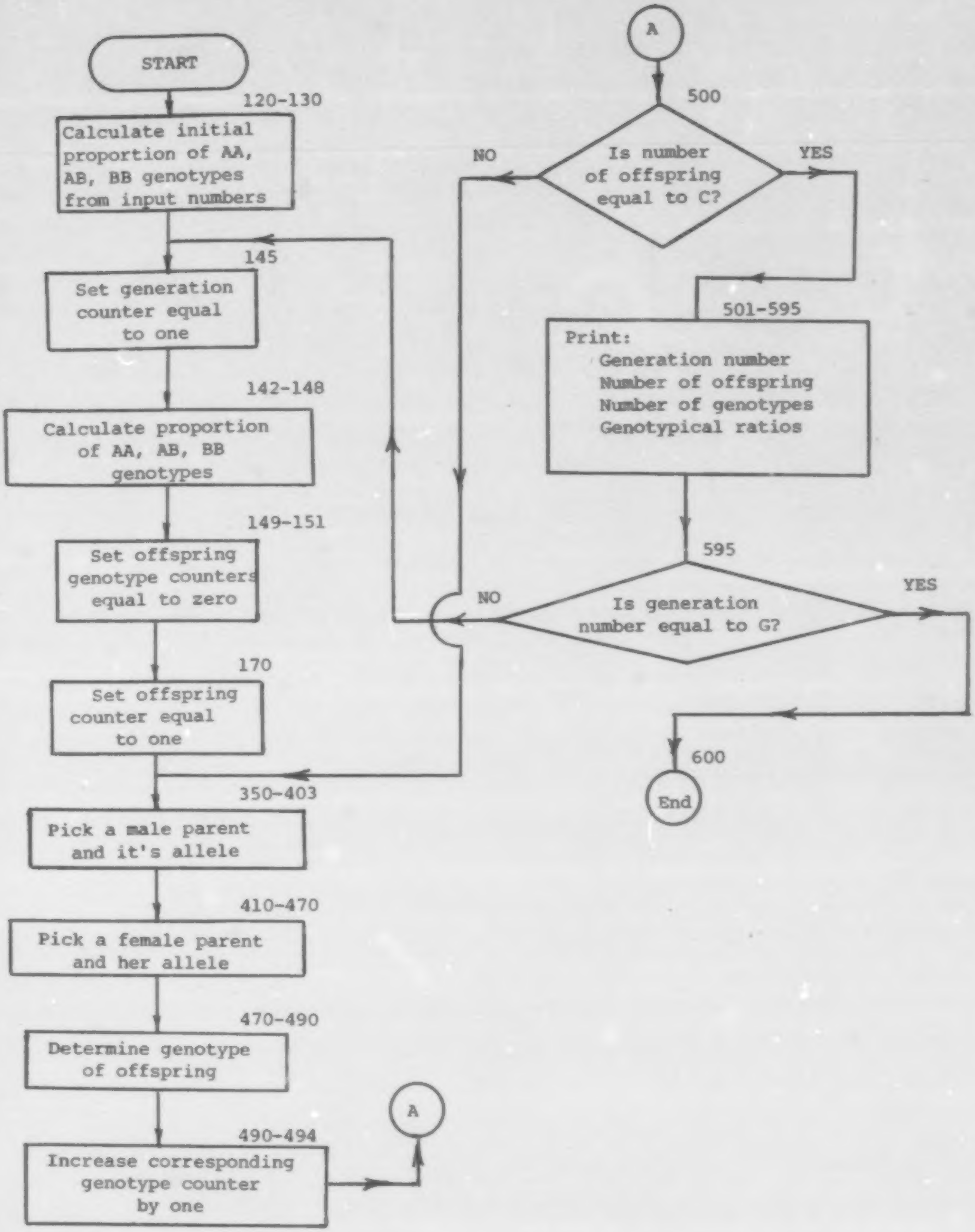
Since the numbers of the respective offspring genotypes are used to calculate the genotypic ratios of the parents of the subsequent generation, it is necessary to grow very many offspring in order to insure a valid estimation of the parental genotypic ratios. A crude method for estimating the required number of generated offspring can be obtained by growing several offspring, calculating their genotypical ratios and then growing several more offspring and again calculating the genotypical ratios of the total number of offspring. If the two sets of ratios are sufficiently close, it is reasonable to assume that sufficient offspring have been grown. (Note all the weasel words). Such a condition has not been included in the program; however, it could be with some programming effort. The problem of the determination of a sufficient number of offspring is similar to the problem of determining a sufficient number of trials when simulating probabilistic phenomena on a computer and is addressed again in the chapter entitled Random Processes.

As previously stated, the simulation is certainly not valid for those populations which are small in number. For such populations, the parental genotypic ratios must be altered after each mating before simulating the growth of the next offspring. Thus, the determination of the genotypic ratios of subsequent generations of small populations requires more computational effort. In addition, the smallness of the population, coupled with the assumption of only a single offspring from each mating (as in a monogamous population), requires a more serious assessment of the implications of the results.

The following paragraphs explain in some detail how the above listed modifications are implemented into the program. A flowchart of the program is shown in figure 4 and the program is listed in figure 5. The number of generations for which the population is to be simulated will be denoted by G, and the number of offspring to be "grown" each generation by C. These numbers are required input.

The modification of our original program to permit the user to specify the original genotypic population is easily accomplished by the insertion of an input statement (line number 114) in which A1, A2 and A3 are the desired original number of AA, AB and BB genotypes respectively. The ability to specify the number of generations G, as well as the number of offspring per generation C, is provided in the same input statement. The program assumes that the same number of offspring are produced each generation; however, the alteration to permit a different number of offspring each generation is slight.

Since the parents of each generation may have a different genotypical constitution, provision must be made to calculate the genotypical ratios anew after each generation in order to properly simulate the random mating of the new parents according to their genotypical ratios. This is accomplished by statements numbered 146-148. The student who has had a course in genetics and is familiar with the well-known Hardy-Weinberg law should realize that the previous statement does not conflict with the conclusion of this law because the law is valid only for very large populations. The symbols G1, G2 and G3 have the same meaning as in the previous program. The symbols A1, A2 and A3 were introduced to signify the original or starting number of


START
120-130
Calculate initial proportion of AA, AB, BB genotypes from input numbers
145
Set generation counter equal to one
142-148
Calculate proportion of AA, AB, BB genotypes
149-151
Set offspring genotype counters equal to zero
170
Set offspring counter equal to one
350-403
Pick a male parent and it's allele
410-470
Pick a female parent and her allele
470-490
Determine genotype of offspring
490-494
Increase corresponding genotype counter by one
A
A
500
Is number of offspring equal to C?
NO
YES
501-595
Print:
Generation number
Number of offspring
Number of genotypes
Genotypical ratios
595
Is generation number equal to G?
NO
YES
600
End

Flowchart to simulate the genotype
description over several generations

Fig. 4

```
GENE1

5 REM         FIRST POPULATION GENETICS PROGRAM
6 REM
7 REM
20 RANDOMIZE
102 PRINT "A1=NO. OF  AA  GENOTYPES IN ORIGINAL POPULATION"
103 PRINT "A2=NO. OF  AB  GENOTYPES IN ORIGINAL POPULATION"
104 PRINT "A3=NO. OF  BB  GENOTYPES IN ORIGINAL POPULATION"
108 PRINT "C=NO. OF OFFSPRING PER GENERATION"
110 PRINT "G=NO. OF GENERATIONS TO RUN PROGRAM"
112 PRINT "TYPE  A1.  A2,  A3.  C,  G."
114 INPUT A1,A2,A3,C,G
115 PRINT
116 PRINT
117 PRINT
118 PRINT "                PROGRAM RESULTS        "
119 PRINT
120 PRINT
125 LET N=A1+A2+A3
128 LET N1=A1/N
129 LET N2=A2/N
130 LET N3=A3/N
131 PRINT "THE INITIAL PROPORTION OF AA GENOTYPES IS";N1
132 PRINT
133 PRINT "THE INITIAL PROPORTION OF AB GENOTYPES IS";N2
134 PRINT
135 PRINT "THE INITIAL PROPORTION OF BB GENOTYPES IS";N3
136 PRINT
139 REM
140 REM  G1, G2, G3 ARE THE NO. OF AA, AB, BB GENOTYPES RESPECTIVELY
141 REM
142 LET G1=A1
143 LET G2=A2
144 LET G3=A3
145 FOR I=1 TO G
146 LET N1=G1/N
147 LET N2=G2/N
148 LET N3=G3/N
150 LET G1=0
151 LET G2=0
154 REM
155 REM  LINES 150-152 SET THE GENOTYPE COUNTER TO ZERO AT BEGINNING
156 REM  OF EACH GENERATION
157 REM
300 REM    LINES 350-463 PICK EACH PARENT AND THEIR ALLELE
310 REM
320 REM     LINES 350-410 PICK A MALE PARENT AND HIS ALLELE
321 REM
340 FOR K=1 TO C
350 LET R=RND
360 IF R<=N1 THEN 403
362 IF N2=0 THEN 400
370 IF R>=(N1+N2) THEN 400
```

Figure 5

```
380 LET R=RND
390 IF R>=.5 THEN 403
400 LET M=0
401 GO TO 410
403 LET M=1
404 GO TO 410
407 REM
408 REM     LINES 410-470 PICK A FEMALE PARENT AND HER ALLELE
409 REM
410 LET R=RND
415 IF R<=N1 THEN 463
417 IF N2=0 THEN 460
420 IF R>=(N1+N2) THEN 460
440 LET R=RND
450 IF R>=.5 THEN 463
460 LET F=0
461 GO TO 470
463 LET F=1
464 REM
465 REM     LINES 470-490 DETERMINE THE GENOTYPE
466 REM
470 IF M+F=2 THEN 492
480 IF M+F=1 THEN 494
490 LET G3=G3+1
491 GO TO 500
492 LET G1=G1+1
493 GO TO 500
494 LET G2=G2+1
501 PRINT
502 PRINT
503 PRINT "GENERAION OF THE OFFSPRING IS";I
505 PRINT
510 PRINT "THE NUMBER OF OFFSPRING IS";C
515 PRINT
520 PRINT "THE NUMBER OF AA GENOTYPES IS";G1
525 PRINT
530 PRINT "THE NUMBER OF AB GENOTYPES IS";G2
535 PRINT
540 PRINT "THE NUMBER OF BB GENOTYPES IS";G3
545 PRINT
550 LET N=G1+G2+G3
555 LET R1=G1/N
558 LET R2=G2/N
560 LET R3=G3/N
570 PRINT
575 PRINT "THE AA GENOTYPIC RATIO IS";R1
580 PRINT
590 PRINT "THE AB GENOTYPIC RATIO IS";R2
595 PRINT
600 PRINT "THE BB GENOTYPIC RATIO IS";R3

READY
```

Figure 5 (continued)

genotypes. Thus, lines 128-130 calculate the genotypic ratios of the intial parent population in order that these ratios may be printed out by lines 131, 133 and 135. By again recalculating the original genotypic ratios in lines 142-148, your author provided a set of statements, lines 146-148, which are used each time the parent genotypic ratios are calculated anew in terms of the numbers of the respective genotypes of the offspring. The generation loop begins at line 145 and separates the calculation of the initial parent genotypic ratios from the calculation of the genotypic ratios for each generation. Lines 150 and 151 set the respective offspring genotype counters to zero and line 340 begins the offspring generation loop.

From the initial number of AA, AB and BB genotypes the initial genotypic ratios are calculated. (Inst. 125 etc.). Since both the male and female parents are assumed to have the aforementioned genotypical distribution, a random number generator is employed to select at random a male parent from such a distribution (Inst. 350-370). If the male parent is an AA or BB, the gamete will carry only an A or B allele respectively. However, if an AB male parent is selected, the fact that it is assumed to be equally likely that the gamete will carry an A or B allele requires that the random number generator again be employed to determine which allele will be transmitted (Inst. 380). The selection of the female parent and her allele is accomplished in an identical manner (Inst. 410-470). The genotype of the resulting offspring is then determined (Inst. 470-490) and the process repeated until the required number of offspring are created. These offspring then become the sole parents of the next generation and the entire process is repeated.

In order that the student may more readily understand the method used to "grow an offspring" we will describe how a mating and resultant determination of the genotype of the offspring is simulated assuming that there are G1 parents who possess AA alleles, G2 parents who possess AB alleles, and G3 parents who possess BB alleles respectively. An example will best illustrate how this is accomplished. Suppose that G1=200, G2=500, and G3=100, i.e. there are 200 parents who are AA genotypes, 500 parents who are AB or BA genotypes and 100 parents who are BB genotypes. Thus, 1/4 of the parental population is AA, 5/8 of the parental population is AB or BA, and the remaining 1/8 is BB. Now, in order to randomly select parents from such a genotypical population, we select a random number between 0 and 1 and determine where this random number lies relative to a linear genotypical description of the population. This determination can perhaps best be described by imagning that the parental genotypical description is "laid out" along a unit interval, OP (see figure 6 below).

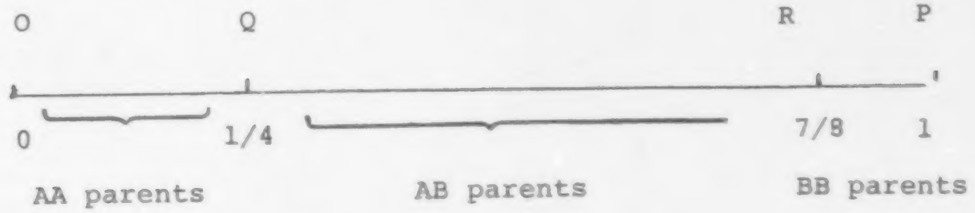


Figure 6

The region OQ represents the portion of parents with the AA genotype, the region QR represents the portion with AB or BA genotype, and the region RP represents the portion with BB genotypes. Since our random number will be represented by some point on OP it will lie in one of the regions OQ, QR or RP. Thus, if the random number is less than 1/4, the male or female parent is said to be AB; and if the number lies between 1/4 and 7/8, the male or female parent is said to be AB or BA; and if the random number is greater than 7/8, the male or female parent is said to be BB. Furthermore, if the parent is determined to be an AA or BB, the allele of the gamete of such a parent will be an A or B allele respectively, and hence there is no need to again utilize the random number generator to determine which allele is forthcoming to the offspring from such a parent. However, if the parent is determined to possess an AB genotype then a random number generator is utilized to determine the allele transmitted to the offspring.

The determination of the allelic pair of the male parent is accomplished by lines 350, 360, 362, and 370 of the program and the corresponding determination for the female parent is made by lines 410, 415, 417, and 420. Lines 380, 390, 400, and 403 determine the allele passed from the male gamete to the offspring and lines 440, 450, 460, and 463 determine the allele given by the gamete of the female parent. The program alteration to permit the growth of several generations is accomplished by lines 145 and 595. The remainder of the program should be easily understood when read in conjunction with the flowchart.

Since the simulation utilizes the notion of male and female parent, we again remind the student that it is assumed that the probability of selection of an AA, AB, or BB genotype is the same for both sexes. We are also assuming equal probability of selecting an A or B allele if either parent is an AB or BA genotype.

The program is very useful for obtaining a quantitative feel for the distribution of genotypes over several generations. By choosing to grow a large number of offspring each generation, that is by setting C equal to a large number, it is possible to study the evolution over successive generations of the genotypic ratios of an "infinite" population. When this is done, the student will note that the ratios calculated after the first generation do not change significantly in successive generations. This is in accord with the famous Hardy-Weinberg law which will be discussed in the following section.

It is evident that such a method of analysis may be expensive of computer time; however, your author again reiterates his purpose in presenting the analysis of quantitative phenomena in the biological sciences in this manner. It is "to enable the student to obtain an appreciation and understanding of quantitative phenomena minimizing a knowledge and use of formal mathematics". In addition, the cost of a computer calculation is decreasing both in time and money, whereas the cost of learning by conventional means is increasing in both time and money. A thorough and complete discussion and analysis of program results would require the use of several statistical ideas. Since such an analysis would require a significant departure from the intent of the work, further discussion of the program results will not be considered.

Figures 7 and 8 illustrate sample output for the program. Your author, for no good reason, chose the initial numbers of genotypes to be 1, 1, 4 and 1, 20, 1. Note that these numbers of genotypes imply initial genotypic ratios of (1/6, 1/6, 4/6) and (1/22, 20/22, 1/22). The student might ask why was not a larger number of parent genotypes chosen? The answer is that there is no need to since it is the genotypic ratios which are important. As stated previously, there are an infinite set of genotypes that can give the same set of genotypic ratios and

```
RUN

GENE1

A1=NO.  OF  AA  GENOTYPES IN ORIGINAL POPULATION
A2=NO.  OF  AB  GENOTYPES IN ORIGINAL POPULATION
A3=NO.  OF  BB  GENOTYPES IN ORIGINAL POPULATION
C=NO. OF OFFSPRING PER GENERATION
G=NO. OF GENERATIONS TO RUN PROGRAM
TYPE  A1.  A2,  A3.  C,  G.
?1, 1, 4, 1000, 5


                PROGRAM RESULTS


THE INITIAL PROPORTION OF AA GENOTYPES IS .166667

THE INITIAL PROPORTION OF AB GENOTYPES IS .166667

THE INITIAL PROPORTION OF BB GENOTYPES IS .666667


GENERATION OF THE OFFSPRING IS 1

THE NUMBER OF OFFSPRING IS 1000

THE NUMBER OF AA GENOTYPES IS 55

THE NUMBER OF AB GENOTYPES IS 364

THE NUMBER OF BB GENOTYPES IS 581

THE AA GENOTYPIC RATIO IS .055

THE AB GENOTYPIC RATIO IS .364

THE BB GENOTYPIC RATIO IS .581


GENERATION OF THE OFFSPRING IS 2

THE NUMBER OF OFFSPRING IS 1000

THE NUMBER OF AA GENOTYPES IS 59

THE NUMBER OF AB GENOTYPES IS 369

THE NUMBER OF BB GENOTYPES IS 572

THE AA GENOTYPIC RATIO IS .059

THE AB GENOTYPIC RATIO IS .369

THE BB GENOTYPIC RATIO IS .572


GENERATION OF THE OFFSPRING IS 3

THE NUMBER OF OFFSPRING IS 1000

THE NUMBER OF AA GENOTYPES IS 69

THE NUMBER OF AB GENOTYPES IS 367

THE NUMBER OF BB GENOTYPES IS 564

THE AA GENOTYPIC RATIO IS .069

THE AB GENOTYPIC RATIO IS .367

THE BB GENOTYPIC RATIO IS .564


GENERATION OF THE OFFSPRING IS 4

THE NUMBER OF OFFSPRING IS 1000

THE NUMBER OF AA GENOTYPES IS 67

THE NUMBER OF AB GENOTYPES IS 378

THE NUMBER OF BB GENOTYPES IS 555

THE AA GENOTYPIC RATIO IS .067

THE AB GENOTYPIC RATIO IS .378

THE BB GENOTYPIC RATIO IS .555
```

Figure 7

```
RUN

GENE1

A1=NO. OF  AA  GENOTYPES IN ORIGINAL POPULATION
A2=NO. OF  AB  GENOTYPES IN ORIGINAL POPULATION
A3=NO. OF  BB  GENOTYPES IN ORIGINAL POPULATION
C=NO. OF OFFSPRING PER GENERATION
G=NO. OF GENERATIONS TO RUN PROGRAM
TYPE  A1.  A2,  A3.  C,  G.
?1, 20, 1, 1000, 5


                PROGRAM RESULTS


THE INITIAL PROPORTION OF AA GENOTYPES IS .0454545

THE INITIAL PROPORTION OF AB GENOTYPES IS .909091

THE INITIAL PROPORTION OF BB GENOTYPES IS .0454545


GENERATION OF THE OFFSPRING IS 1      GENERATION OF THE OFFSPRING IS 3

THE NUMBER OF OFFSPRING IS 1000       THE NUMBER OF OFFSPRING IS 1000

THE NUMBER OF AA GENOTYPES IS 266     THE NUMBER OF AA GENOTYPES IS 228

THE NUMBER OF AB GENOTYPES IS 486     THE NUMBER OF AB GENOTYPES IS 512

THE NUMBER OF BB GENOTYPES IS 248     THE NUMBER OF BB GENOTYPES IS 260

THE AA GENOTYPIC RATIO IS .266        THE AA GENOTYPIC RATIO IS .228

THE AB GENOTYPIC RATIO IS .486        THE AB GENOTYPIC RATIO IS .512

THE BB GENOTYPIC RATIO IS .248        THE BB GENOTYPIC RATIO IS .26


GENERATION OF THE OFFSPRING IS 2      GENERATION OF THE OFFSPRING IS 4

THE NUMBER OF OFFSPRING IS 1000       THE NUMBER OF OFFSPRING IS 1000

THE NUMBER OF AA GENOTYPES IS 235     THE NUMBER OF AA GENOTYPES IS 232

THE NUMBER OF AB GENOTYPES IS 505     THE NUMBER OF AB GENOTYPES IS 473

THE NUMBER OF BB GENOTYPES IS 260     THE NUMBER OF BB GENOTYPES IS 295

THE AA GENOTYPIC RATIO IS .235        THE AA GENOTYPIC RATIO IS .232

THE AB GENOTYPIC RATIO IS .505        THE AB GENOTYPIC RATIO IS .473

THE BB GENOTYPIC RATIO IS .26         THE BB GENOTYPIC RATIO IS .295
```

Figure 8

your author chose a simple set. The 1000 offspring were chosen to be grown each generation because this number represented a happy medium between a very large number of offspring requiring much more computer time and a very small number of offspring which would not have been representative of an infinite population.

An analysis of the results of both runs reveals that the genotypic ratios for the first generation offspring is considerably different than the genotypic ratios of their parents. However, there is not a significant difference, from generation to generation, of the genotype ratios of the subsequent generations. As stated above, this observation is in accord with the Hardy-Weinberg principle. The student is urged to experiment with different initial genotypic ratios and offspring calculation. By so doing, a good feel for some of the principal results of classical population genetics can be obtained.

The Hardy-Weinberg Principle

The observation that the genotypic ratios of the offspring did not significantly change after the first generation, suggests that such behavior may indeed be the case regardless of the distribution of the initial genotypic ratios. Further confirmation of this hypothesis may be obtained by making other runs with different initial genotypic ratios and using a larger number of offspring per generation. The hypothesis that the genotypic ratios do not change after the first generation in an evolving infinite population in which pure random mating occurs is the thesis of the Hardy-Weinberg law. It is customary in genetics texts to demonstrate the validity of the Hardy-Weinberg law using only first year high school algebra. The ability to demonstrate, independent of any specific numerical values, the validity of a properly specified assertion, is one of the significant advantages of mathematics. Nevertheless, it is necessary to "discover" or "find" assertions worth verifying. This may be done with the aid of the language of mathematics, but is usually done by observation of empirical and/or computer generated results. In this work, the latter viewpoint is emphasized.

In order to gain more confidence in our suggested hypothesis, we will examine in detail a particular case. We will choose a particular set of numerical values for the numbers of initial genotypes and also choose a given number (large) of offspring per generation and carry through all of the gory arithmetic in a manner as close as possible as is done by the computer. Of course, such a procedure proves absolutely nothing (except to disprove the assertion, in the event the results of the arithmetic calculations contradict the suggested hypothesis). However, such a procedure usually does provide insight into the entire process. Moreover, if the results of the arithmetic calculation are in agreement with the assertion, the evidence for the validity of the assertion is increased, and it may then be worthwhile to attempt to verify the assertion with the aid of the language of mathematics. The actual act of performing the arithmetic calculation is frequently an aid in establishing or developing the necessary mathematical tricks. It should be noted that the computer program relies on the simulation of random processes with the aid of a random number generator. We shall assume that random mating is indeed simulated by such a process and we shall assume that the selecting of a large number of offspring is equivalent to growing an infinite populations. In carrying out the arithmetic calculations, use will be made of the frequency interpretation of probability.

In the computer program, it was assumed that the number of AA, AB and BB genotypes was the same for the males as it was for the females. Thus, the genotypic distribution of both parents may be given by specifying the number of AA, AB and BB genotypes respectively. For this particular problem we suppose that the initial parent population consists of 100,000 pairs of males and females with the following distribution of genotypes: 30,000AA; 50,000AB; and 20,000BB. Hence, the initial parent genotypical ratios are 3/10 AA; 5/10 AB; and 2/10 BB. Because it is assumed that there is an equal probability of a male or a female being born, there will be an equal number of males and females in the offspring population and each sex will have the same number of AA, AB and BB genotypes respectively. The genotypical ratios of these 100,000 offspring will then be the genotypic ratios of the 100,000 males and 100,000 females that will act as parents for the next generation. This assumes that each sex has the same respective genotypic ratios.

The calculation for the determination of the number of each genotype in the offspring populations is done in the following way. Since random mating in an infinite population is assumed, the proportion of each type of mating, AA male x AB female, BB male x AA female, etc., will depend upon the product of the proportion of the male genotype and the proportion of the female genotype. This follows from the fact that when an individual chooses a mate, the likelihood of that mate being a particular genotype is proportional to the number of that particular genotype in the partner's population. Hence, for the genotypic populations assumed in this example, any male has a 3/10 chance of mating with an AA female, a 5/10 chance of mating with an AB female and a 2/10 chance of selecting a BB female mate. Since there is a 3/10 chance that the mating male will be an AA genotype, the likelihood, out of all possible matings, of an AA male mating with an AA female partner is 3/10 x 3/10 or 9/100. Thus, random mating implies that the number of matings of any specified pair of genotypes is proportional to the product of the proportions of the genotypes in the population of each partner. Hence, we can write that the proportion of:

AA males x AA females is 3/10 x 3/10 or 9/100,
AA males x AB females is 3/10 x 5/10 or 15/100,
AA males x BB females is 3/10 x 2/10 or 6/100,
AB males x AA females is 5/10 x 3/10 or 15/100,
AB males x AB females is 5/10 x 5/10 or 25/100,
AB males x BB females is 5/10 x 2/10 or 10/100,
BB males x AA females is 2/10 x 3/10 or 6/100,
BB males x AB females is 2/10 x 5/10 or 10/100, and
BB males x BB females is 2/10 x 2/10 or 4/100.

Frequently, such a listing is presented in array form as:

		MALES		
		AA (3/10)	AB (5/10)	BB (2/10)
FEMALES	AA (3/10)	AA x AA = 3/10 x 3/10 or 9/100	AB x AA = 5/10 x 3/10 or 15/100	BB x AA = 2/10 x 3/10 or 6/100
	AB (5/10)	AA x AB = 3/10 x 5/10 or 15/100	AB x AB = 5/10 x 5/10 or 25/100	BB x AB = 2/10 x 5/10 or 10/100
	BB (2/10)	AA x BB = 3/10 x 2/10 or 6/100	AB x BB = 5/10 x 2/10 or 10/100	BB x BB = 2/10 x 2/10 or 4/100

The student should note that the computer program, when randomly selecting mates, did not alter the proportions of the remaining genotypes as the partners were selected. This was done because it was assumed that the parent populations were infinite. By now the student should recognize the importance, for infinite populations, of the ratio of the genotypes, not the absoulute number of the genotypes. In fact, for infinite populations, it makes no sense whatsoever to talk about numbers of genotypes. However, in the program the number of each genotype in the offspring population determines the genotypic ratios for the parent generation. This method of determining the genotypic ratios is possible because the computer generates only a finite number of matings; it cannot generate an infinite number of matings because of its finite capacity. Thus, the only purpose in growing a large number, say 100,000, of offspring is to assure that the computer grown genotypic distribution approximates the genotypical distribution resulting from the growth of an infinite number of offspring.

The determination of the number of AA, AB or BB offspring genotypes is made by first determining the respective number of parent crosses and then, in conjunction with the number of offspring, determining the number of the genotypes resulting from each type of crossing. We work this out in complete detail. The respective number of parent crosses are obtained by multiplying the proportionate number of crosses with the total number of offspring. It is assumed that one cross results in one offspring. The numbers are:

for AA males x AA females, 9/100 x 100,000 or 9,000,
for AA males x AB females, 15/100 x 100,000 or 15,000,
for AA males x BB females, 6/100 x 100,000 or 6,000,
for AB males x AA females, 15/100 x 100,000 or 15,000,
for AB males x AB females, 25/100 x 100,000 or 25,000,
for AB males x BB females, 10/100 x 100,000 or 10,000,
for BB males x AA females, 6/100 x 100,000 or 6,000,
for BB males x AB females, 10/100 x 100,000 or 10,000, and
for BB males x BB females, 4/100 x 100,000 or 4,000.

Since it is assumed that there is no sexual discrimination acting in favor of or against one or the other of the alleles that is passed from the parent to the offspring, the likelihood of A or B allele passed by an AB parent is the same regardless of the sex of the parent. The numbers of the genotypes produced by each cross is then seen to be:

from the AA male x AA female, 9,000 AA, since only an AA offspring can be produced;
from the AA male x AB female, 7,500 AA and 7,500 AB, since one-half of the offspring are AA and the other half are AB;
from the AA male x BB female, 6,000 AB, since only an AB offspring can be produced;
from the AB male x AA female, 7,500 AA and 7,500 AB, since one-half the offspring are AA and the other half are AB;
from the AB male x AB female, 6,250 AA, 12,500 AB and 6,250 BB, since one-fourth of the offspring are AA and BB and the remainder are AB;
from the AB male x BB female, 5,000 AB and 5,000 BB, since one-half of the crossing produce AB offspring and the other half produce BB offspring;

from the BB male x AA female, 6,000 BA, since only a BA can be produced;

from the BB male x AB female, 5,000 BA and 5,000 BB, since one-half are BA and the other half are BB, and finally

from the BB male x BB female, 4,000 BB, since only BB offspring can be produced from such a mating.

With the aid of the assumption that an AB genotype is identical to a BA genotype, that is there is no sex preference for the origin of a particular allele, the numbers of AB and BA genotypes may be combined. Thus, the numbers of offspring genotypes are: 30,250 AA, 49,500AB and 20,250BB. Note that there are 2x30,250 + 49,500 or 110,000 A alleles and 2x20,250 + 49,500 or 90,000 B alleles. Hence, the total number of alleles has been preserved. This means that there has been no immigration nor emigration of either type of genotype nor has there been any mutation during the production of the offspring. The genotypic ratios are: 3025/10000 for AA, 4950/10000 for AB and 2025/10000 for BB. The numbers of genotypes, as well as the corresponding genotypic ratios, for the offspring are different than those assigned initially to their parents. This difference in the ratios is in agreement with the computer results. The computer results also indicated that, when these progeny acted as parents, the respective numbers of these genotypes of their offspring would be equal to the numbers of the genotypes of the parents. Thus, there would be no further change in the genotypic ratios of subsequent offspring generations. To check this observation, we proceed to again calculate the numbers of genotypes, together with their genotypic ratios, of the offspring of these progeny. Since the first offspring become parents to the new progeny, we use the genotypic ratios just calculated as the genotypic ratios of the new parents.

The calculation proceeds just as before and so some of the calculations are combined in order to shorten the presentation. It follows that:

the number of AA male and AA female crosses is
3025/10000 x 3025/10000 x 100,000 or 9150.625,

the number of AA male and AB female crosses is
3025/10000 x 4950/10000 x 100,000 or 14973.75,

the number of AA male and BB female crosses is
3025/10000 x 2025/10000 x 100,000 or 6125.625,

the number of AB male and AA female crosses is
4950/10000 x 3025/10000 x 100,000 or 14973.75,

the number of AB male and AB female crosses is
4950/10000 x 4950/10000 x 100,000 or 24502.5,

the number of AB male and BB female crosses is
4950/10000 x 2025/10000 x 100,000 or 10023.75,

the number of BB male and AA female crosses is
2025/10000 x 3025/10000 x 100,000 or 6125.625,

the number of BB male and AB female crosses is
2025/10000 x 4950/10000 x 100,000 or 10023.75, and finally

the number of BB male and BB female crosses is
2025/10000 x 2025/10000 x 100,000 or 4100.625.

Hence, the numbers of genotypes produced by these respective matings is:

from the AA male x AA female - 9150.625 AA,
from the AA male x AB female - 7486.875 AA and 7486.875 AB,
from the AA male x BB female - 6125.625 AB,
from the AB male x AA female - 7486.875 AA and 7486.875 AB,
from the AB male x AB female - 6125.625 AA, 12251.25 AB and 6125.625 BB,
from the AB male x BB female - 5011.875 AB and 5011.875 BB,
from the BB male x AA female - 6125.625 AB,
from the BB male x AB female - 5011.875 AB and 5011.875 BB,
and finally from the BB male x BB female - 4100.625 BB.

Thus, the numbers of each offspring genotype produced in the second generation are: 30,250 AA, 49,500 AB, and 20,250 BB. These numbers are the same as that grown in the first generation and consequently this particular numerical example reinforces the credence of the hypotheses that the genotypic ratios remain constant after the first generation. Of course, this assertion is subject to all of the hypotheses and assumptions used in carrying out the numerical example. We reiterate, the purpose in carrying out all of the gory arithmetic in such detail was to clearly indicate just where all of the assumptions entered the development. It must again be emphasized that such a numerical calculation in no way constitutes a proof of the assertion. A proof would require that the initial genotypic ratios be specified independent of any particular numerical values and then the argument carried through with the aid of elementary algebra. Because the algebraic proof requires the specification of the genotypic ratios rather than the numbers of each type of genotype, the mathematical proof proceeds somewhat differently than has been indicated above. In the preceding calculation, the results were carried out to decimal fractions just to indicate that the results are indeed numerically accurate for this particular set of initial numbers of genotypes. The student is urged to choose another set of numbers and to then carry out the calculation in order to more thoroughly understand the process and to see how the assumptions enter the calculation.

Your author apologizes for the arithmetic detail of the previous work; however, it has been his experience that many non-mathematically oriented students have difficulty following the terse mathematical arguments presented in the language of mathematics. It is especially difficult for such students to understand where and how the basic assumptions enter the development.

To summarize, we present a formal statement of the Hardy-Weinberg law. The law states that the proportions of the alleles at a particular locus, as well as the proportion of the genotypes derived from them, remain constant from generation to generation. Furthermore, these genotypic proportions will be found after the first generation of mating and are independent of the genotypic proportions of the original parental population. The conditions under which the law is valid are:

1. The population must be sufficiently large so that chance changes in gene ratios are insignificant.
2. Mutation must either not occur or must have already reached its own equilibrium.
3. Reproduction must be random, that is, the chance of mating is independent of the genotype.
4. All types of matings are equally successful in producing offspring, i.e. there is no genotypical or sexual survivability dependence.
5. There is no preference by sex of the transference of an allele from the parent to the offspring.
6. The allelic ratios of both sexes are the same.

In all of the preceding developments, the terms 'ratio' and 'proportion' have been used interchangeably; however, most works in genetics use the term 'frequency'. The terms ratio and proportion were deliberately used because, to most students, these terms do indeed denote a comparison of the relative magnitudes of two quantities, whereas the term frequency usually refers to the number of occurrences of an event. The student should note this distinction in the use of these terms in order that he or she not be confused when consulting works in genetics. (In these works, the word frequency is taken to mean proportion or ratio).

The constancy of the genotypic ratios for all succeeding generations after the first, is referred to as Hardy-Weinberg equilibrium. Since a great many populations closely approximate the assumptions required for the Hardy-Weinberg law to obtain, the equilibrium conclusion can be used to study the effects of various exterior influences, such as migration, mutation, etc., on the genetic evolution of the population. These effects are imposed on the population, the resultant genotypic ratios are calculated for subsequent generations, and then compared to those that would be inferred by the Hardy-Weinberg law. This provides a yardstick by which the effects of exterior induced genetic alterations may be measured.

The equilibrium conclusion of the law, enables the prediction of future genetic behavior of populations which closely approximate the assumptions in the law. There are thus two different, but important, implications of the law.

If the population is in Hardy-Weinberg equilibrium, that is, the genotypic ratios are not changing from generation to generation, there is a very simple relation between the genotypic ratios and the allelic ratios. In the preceding example, after the initial offspring were grown, there were 30,250 AA, 49,500 AB, and 20,250 BB genotypes. Thus, there were 55,000 A alleles and 45,000 B alleles. In terms of proportions, these numbers are respectively, 0.3025 AA, 0.495 AB and 0.2025 BB, and 0.55 A alleles and 0.45 B alleles. Now, $.3025 = (.55)^2$, $.495 = 2x.55x.45$ and $.2025 = (.45)^2$. This suggests that if a denotes the proportion of A alleles, and b denotes the proportion of B alleles, that the proportion of AA genotypes is a^2, the proportion of AB genotypes is 2ab and that the proportion of BB genotypes is b^2. This can be shown to be true for all proportions of A and B alleles providing the population is in Hardy-Weinberg equilibrium and the sum of the two allele proportions is one. These relations are very useful in analyzing single loci genetic data for populations approximating the Hardy-Weinberg hypotheses. The relation, $a + b = 1$ is also useful in deriving relationships among the various genotypic and allelic ratios.

As stated at the beginning of this section, by making other runs with different initial conditions, further confirmation of the Hardy-Weinberg law may be obtained. Figures 9 and 10 graphically portray results obtained by running the program for 30 generations. In figure 9, the initial respective numbers of parent AA, AB and BB genotypes were 1, 1, and 4, and in figure 10 the initial numbers were 1, 20, and 1.

The results show that, for a growth of 1000 offspring per generation, the BB genotypic ratio remains fairly constant after the first generation. There is an apparent upward drift as the number of generations increases. This is probably due to the lack of complete randomness of the random number generator or it may be an example of mild genetic drift; genetic drift is defined in the next section. The evolution of the BB genotypic ratio for the run corresponding to 100 offspring per generation also tends to confirm the Hardy-Weinberg principle. However, these results fluctuate far more than the results obtained using a growth of 1000 offspring. An even wider fluctuation is indicated by the results wherein only 10 offspring are grown each generation. These observations concerning the fluctuation of the population reinforce the earlier observation that the variability in the results increases with a decrease in the population. Such fluctuations in results are characteristic of all Monte-Carlo type calculations in which the random sample is small.

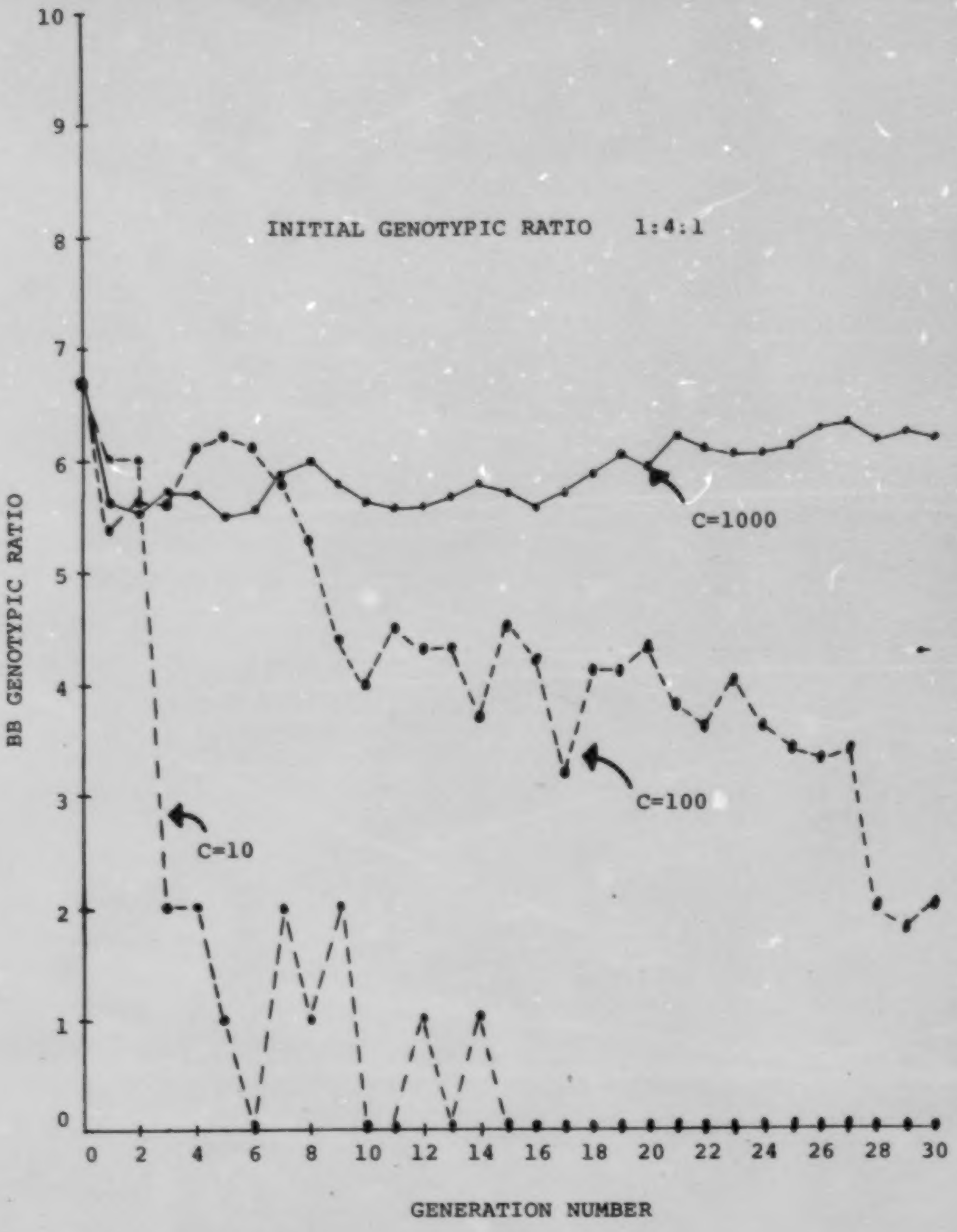

INITIAL GENOTYPIC RATIO 1:4:1
BB GENOTYPIC RATIO
10
9
8
7
6
5
4
3
2
1
0
C=1000
C=100
C=10
0
2
4
6
8
10
12
14
16
18
20
22
24
26
28
30
GENERATION NUMBER

Fig. 9

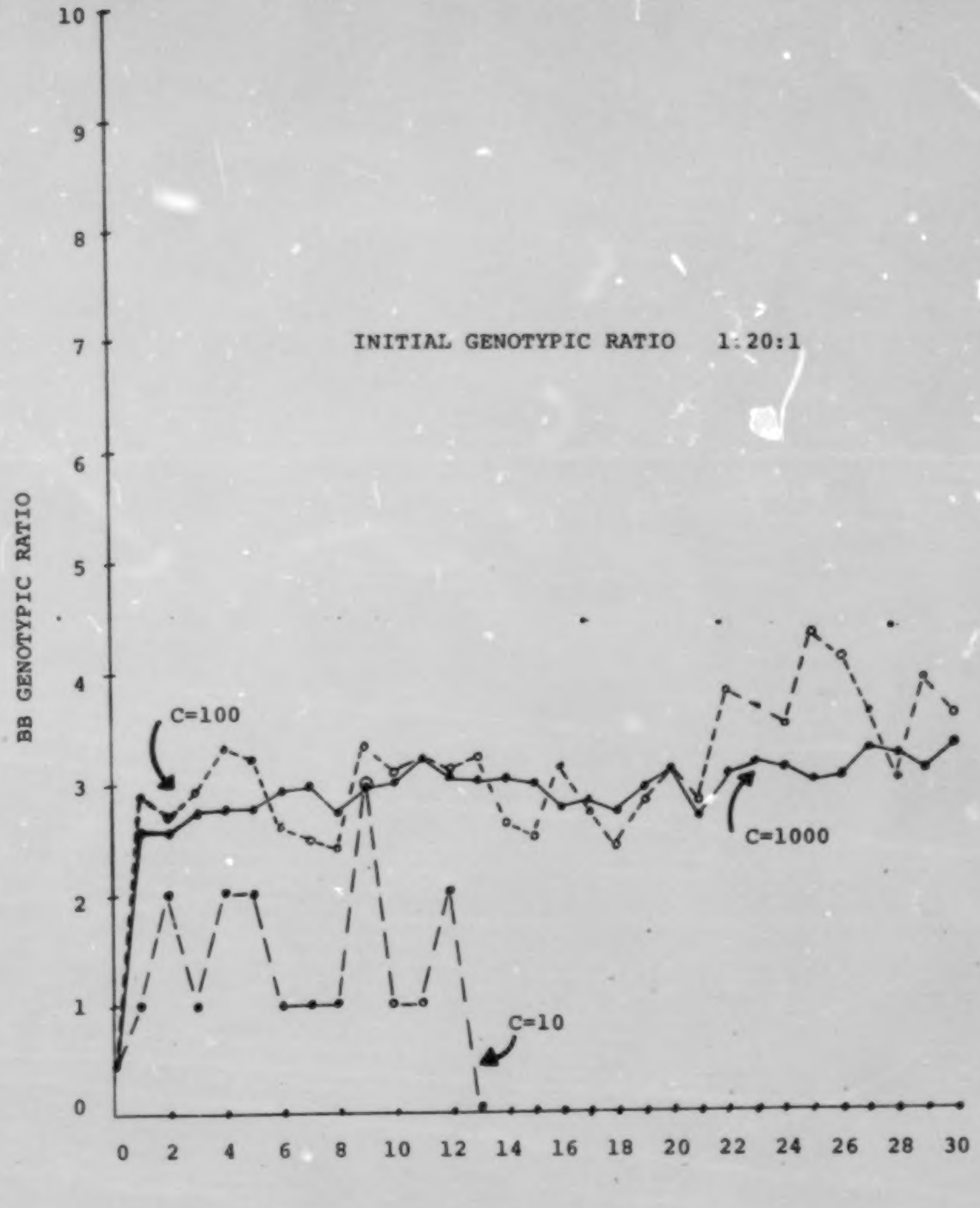

INITIAL GENOTYPIC RATIO 1:20:1
BB GENOTYPIC RATIO
GENERATION NUMBER
C=100
C=1000
C=10
0
1
2
3
4
5
6
7
8
9
10
0
2
4
6
8
10
12
14
16
18
20
22
24
26
28
30

Fig. 10

Gene Dispersal

Both runs labeled C=10, reveal the startling behavior that the BB genotypic ratio reaches zero and eventually stays there. The fact that the BB genotypic ratio remains zero implies that the AB genotypic ratio must also become and remain zero. If such were not the case, the equal probability of an A or a B allele being transferred from parent to offspring would sooner or later have resulted in a mating of a female AB and a male AB genotype; and such a mating would have produced a BB genotype at least one-fourth of the time. An examination of the computer runs revealed that no AB genotypes were in existence after the BB genotypic ratio had remained zero for a sufficient number of generations and the population was made up of only AA genotypes.

The phenomena of an initial distribution of parent alleles consisting of both A and B alleles evolving into a population in which only one allele is present is called gene fixation. No further change is possible in the genetic make-up of the population once the fraction of either the A or the B alleles becomes 0. In the two runs discussed above, the AA genotypic ratio becomes fixed at unity. In the event that the allelic ratio becomes unity, the population is said to be homozygous for that locus. Such a population is called homoallelic. As the results have indicated, gene fixation increases as the population decreases. Thus, small populations exhibit decreased heterozygosity and a loss of variability.

If the behavior of the AA genotypes over successive generations is considered, it will be seen that the AA genotypic ratios also fluctuate from generation to generation. The fluctuation increases with decreasing population. The phenomena of the fluctuation of genotypic ratios due to random alterations is called genetic drift. As is noted from the previous discussion, the notion of gene fixation and the process of genetic drift are closely related.

If the run corresponding to 10 offspring per generation had been repeated, the result could equally well have been that eventually the AA genotype would have disappeared from the gene pool. Thus, the AA genotype might have become and remained zero. Further runs would have yielded different results such as either the A or the B allele disappearing from the population in a greater, or in a smaller, number of generations. Other results would indicate a greater, or a lesser, variation in the BB genotypic ratio. Nevertheless, because of the random nature of the process and the finite size of the population, eventually the population will become homoallelic; that is, will consist only of A alleles or only of B alleles.

It is to be emphasized that, since the genotypic ratios were not changed as each of the 10 offspring were grown, the simulation has mimiced the random mating of 10 pairs of individuals out of an infinite number of such pairs having the same genotypical distribution. Thus, it is assumed that the genotypic distribution of the 10 mating pairs is the genotypic distribution of the infinite population. This means that the genotypic distribution of the infinite population changes in accord with the change in the genotypic distribution as calculated from a sample of only 10 offspring. This is a rather unrealistic presumption and consequently, its implications will not be pursued. Instead, we will use the interpretation that the generation to generation variation of the genotypic distribution of a small mating population can be discussed assuming completely random mating with no change in the parent genotypic ratios as the offspring are grown. In the next section, a more realistic model of the evolution of a finite population will be used. We now continue the discussion of the genotypic variation by examining its variation in small populations.

The number of generations necessary for the population to become homoallelic increases as the sample or offspring population increases. This is verified by the results shown in figures 9 and 10. These results show that by 20 generations or so, one or the other of the alleles had completely disappeared from the population. In contrast, sample populations consisting of 100 or 1000 offspring per generation exhibited relatively small variation in the BB genotypic ratios, let alone indicating a disappearance of one or the other of the alleles by 30 generations.

Since there is such variation from run to run, it is of interest to make several runs and to then construct histograms for the distribution of the number of AA, AB and BB offspring respectively. The histograms should better indicate the variation possible from run to run. In this way it may be possible to study the variation in genotypic ratio with the number of generations. The variation of the genotypic ratio over several generations is called gene dispersal.

One method of examining this variation is to make several runs, each of G generations in duration, and for each run, and for each generation of the run, record the number of BB genotypes among the offspring. Then, for a given generation, form the histogram consisting of the total number of times, for all of the runs, 10 BB genotypes were recorded, the total number of times 9 BB genotypes were recorded, etc. down to the total number of times zero BB genotypes were recorded. If histograms are made for successive generations, say generations 1, 5, 10, 20, 50 and 100, the resultant set of histograms gives a graphical portrayal of the dispersal of the BB genotype.

Since the number of A alleles is given by the sum of the number of the AA genotypes, plus one-half the number of AB genotypes, it is easy to also record the number of A alleles for the offspring of each generation. In a similar manner, the number of B alleles may also be recorded. Schaffer, page 52, ref. 1, displayed the phenomena of gene dispersal by portraying the variation, from generation to generation, of the ratio of the number, N_A, of A alleles to the number, N_B, of B alleles. His computer generated results were obtained from a mimicing of the growth of 400 populations over 32 generations, each population having 8 matings per generation and an initial distribution of 2 AA, 4 AB, and 2 BB individuals. In order that the student may better appreciate the significance of the ratio N_A/N_B, we note that if $N_A/N_B = 0$, the number of A alleles is 0 and the number of B alleles is 16. On the other hand, if N_A/N_B is infinite, the number of A alleles is 16 and the number of B alleles is 0. Also, $N_A/N_B = 6/10$ implies that there are 6 A alleles and 10 B alleles and $N_A/N_B = 14/2$ means that there are 14 A alleles and 2 B alleles. Schaffer recorded, for each of the 32 generations, each of the possible 17 values that N_A/N_B could assume. For a selected set of successive generation numbers, numbers 1, 2, 4, 8, 16 and 32, the frequency of the possible values that could be assumed by N_A/N_B, were plotted against the ratio number. These numbers ranged from 0 to 16 with number 0 corresponding to the ratio, $N_A/N_B = 0$, and the number 16 corresponding to the ratio, N_A/N_B = infinite. In this way, Schaffer constructed a sequence of six histograms which clearly indicated the increasing dispersal with increasing generations.

Related Projects

As a suggested project, the student is urged to modify the preceding program to print out the information necessary to enable the drawing of the histograms. If a plotter is available, a set of computer drawn histograms can be generated. By varying the initial parent genotypic ratios, it is possible to study the variation in spread after a specified number of generations due to different initial genotypic ratios. It is also possible to study the effect of offspring population size on the variation in the spread of the genotypic ratio. The modification of the program to obtain the necessary information is tedious but rather straightforward. The student should note that an analysis of such variation is complicated because it is rather difficult to decide on what is a reasonable measure of such variation. The fact that the population became homoallelic after a number of generations, suggests that the average number of generations required for the population to become homoallelic may be a reasonable measure of the dispersal due to random mating in small sample populations. The necessary program alteration to discuss this is straightforward.

Small Population

This section describes the development of a computer program to mimic the genotypical evolution of a small population. The population is assumed to consist of males and females who mate in a monogamous manner. It is again assumed that the offspring will be the parents of the next generation so that the genotypical distribution of the offspring is assumed to be the genotypical distribution of the parents for the subsequent generation. The basic ideas are straightforward and rather easy to understand; however, their implementation in a program requires some attention to detail. In the discussion to follow, frequent reference will be made to specific sets of lines of the program. In this way, the discussion can be more readily related to the program which is presented in figure 18 (p. 7.76). As in the previous genetics programs, the program will be restricted to describing the genetic evolution in terms of 2 alleles at a single locus.

As previously stated, in a large or infinite population, the probability of selecting an AA, AB or BB genotype male parent remains the same no matter how many AA, AB or BB male parents have been previously selected. Thus, when repeatedly selecting parent genotypes, the parent genotypic ratios do not have to be changed after each selection. Such is not the case, however, for small mating populations. When the population is small, say less than 25 or 50, it is necessary to alter the parent genotypic ratios in accord with the number of previous matings.

For example, suppose the starting population was composed of 2AA, 4AB and 3BB male parent genotypes. The initial male genotypic ratios would then be 2/9AA, 4/9AB and 3/9BB. Now suppose it is desired to investigate the offspring resulting from the mating of these males with the females in the population. If the first two males which mate are AA, then there are no other AA male parents in the parent population. Consequently, the male parent genotypic ratios are 0AA, 4/7AB and 3/7BB. In contrast, if the genotypic numbers 2AA, 4AB and 3BB characterized the genotypical make-up of an infinite population, after having selected 2AA male parents, the ratios of male parent genotypes would be 2/9AA, 4/9AB and 3/9BB. These ratios are the same as the original genotypic ratios. The important fact is that for infinite populations, the genotypic ratios of the parents do not change after a finite number of matings. In contrast, for small populations, the genotypic ratios change after every mating. Thus, in mimicing the growth of a small population, the genotypic ratios must be properly altered after each mating.

In order to clarify the procedure for altering the genotypic ratios, some examples will be presented. The student should note that, for the example considered in the previous paragraph, the probability of selecting an AA male as the first of the male parents to mate is 2/9. However, after this AA male has been selected, there is only one AA male remaining in the parent population, and the number of male parents has been reduced to 8. Thus, the probability of selecting a second male AA parent is 1/8, the probability of selecting a male AB is 4/8, and of selecting a male BB parent is 2/8. Hence, after a male AA parent has been selected, not only does the probability of selecting an AA male parent change, but so does the probability of selecting either an AB or a BB male parent. As a further example, suppose an AA male had been selected first, a BB male next and then another BB male parent selected. The genotypic numbers are then 1AA, 4AB and 1AB and hence, the genotypic ratios are 1/6, 4/6, and 1/6 respectively. In this case, the probability of selecting an AA male parent for the next mating is 1/6, and for the respective selection of an AB and a BB male parent, the probability is 4/6 and 1/6.

This rather lengthy discussion was given to better enable the student to "see" how the probability of selecting a parent changes in accord with the number of genotypes of the parents that have previously mated during that generation. The genotypical ratios for the selection of female parents must change in the same manner. A procedure for effecting such a change is described below and is used in the program in lines 3040 to 3180 for the male parents and in lines 3500 to 3640 for the female parents.

In terms of the notation used in the program, the alteration of the genetic ratios is accomplished in the following way. Let N1, N2 and N3 denote the respective initial number of AA, AB and BB male genotypes and let T1 = N1+N2+N3. The initial genotypic ratios are N1/T1, N2/T1 and N3/T1. Now let U1, U2 and U3 denote the number of AA, AB and BB male parents that so far have mated in this generation. The genotypic ratios of the male parents after U1+U2+U3 matings are respectively:

(N1-U1)/(T1-U1-U2-U3), (N2-U2)/(T1-U1-U2-U3) and (N3-U3)/(T1-U1-U2-U3).

The number of male matings of each genotype prior to the present mating are counted by the counters U1, U2 and U3 and the number of each such female genotype matings is counted by the counters U4, U5 and U6. The appropriate counter U1, U2 or U3, is increased by one each time the respective male parent genotype is selected to mate. The appropriate counter U4, U5 or U6, is also increased by one each time the respective female parent genotype is selected to mate. In this way, the counters are changed after each mating and consequently, so also are the genotypic ratios changed in the desired way. Of course, at the beginning of the next generation, before any offspring are grown, these counters must all be initialized to zero. (See lines 5550 to 5670).

The technique of randomly selecting a male or female parent from such a distribution of parent genotypes is the same as that used in the previous program. For the male parents, the genotypic ratios are "layed out" on the interval (0,1) as shown in figure 11.

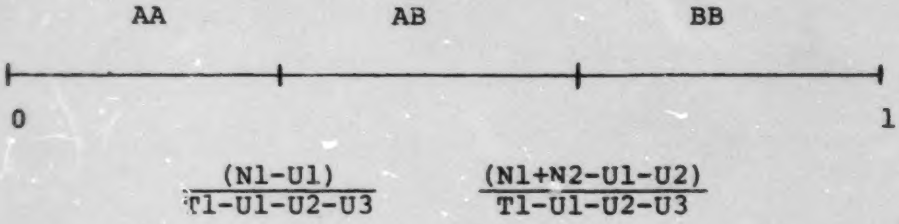


Figure 11

The lengths of the intervals are equal to the numerical value of the genotypic ratios. In accord with the previous discussion, the interval boundaries are changed after every mating to adjust to the new genotypic ratios. A male parent genotype is chosen by using the random number generator to select a random number and the interval which contains the random number defines the genotype of the parent. When U1=N1, all male AA parent genotypes have mated and whenever U2=N2, the boundary points of the AB genotype interval coincide. Consequently, the AB interval has vanished and in this event there are no more AB genotypes in the parent population. Similar remarks hold about the BB genotype interval. This method of determining the genotypic ratios assures that no more than the original number of a specified male parent genotype can be selected. Analogous statements apply to the selection of the female parent genotype.

For a small population, the evolution of the genotypic ratios over several operations will be mimiced by writing a program which:

(a) accepts as inputs the respective number of male and female AA, AB and BB genotypes,

(b) varies the probability of selecting the parents in accord with the number of AA, AB or BB genotypes that have previously been selected, then calculates the genotypes of the offspring,

(c) tabulates the respective number of AA, AB or BB offspring genotype for each generation, and

(d) repeats the above process for several generations.

The student should note that, by growing a large number of offspring each generation, the program could also be used to mimic the genetic evolution of large monogamous populations. However, such a use would be quite expensive of computer time.

In the program, provision will be made to specify as input, the number of male and the number of female parent AA, AB and BB genotypes. These will be denoted by N1, N2, N3 and N4, N5, N6 respectively. 01, 02, 03 and 04, 05, 06 will denote the AA, AB and BB male and female offspring genotypes respectively. Because the mating is assumed to be monogamous, the assumption of a single offspring per mating would result in a halving of the population each generation. Hence, provision will be made to produce a random number of offspring from each mating, which number will result, over several matings and generations, in a specified average number of offspring per mating. The selected average will be two offspring, but any other average can be ensured by a simple alteration of the program. (See lines 500 and 510).

In the program to be described, it will be assumed that, if more than one offspring results from a single mating, each of the offspring will have the same genotype. In other words, the offspring from a single mating will be assumed to be genetically identical. In the event more than one offspring results from a mating, the sex of all such offspring will be determined in a random manner. (See lines 4050 to 4080, lines 4230 to 4270, and lines 4420 to 4460). This means it will be possible for there to be born, in a single generation, an unequal number of males and females. Since monogamous mating is assumed, and the offspring of the previous generation are assumed to become the parents of the subsequent generation, the number of matings in the new generation will be the minimum number of male and female offspring from the previous generation. For the initial mating population, this determination is made in lines 1750 to 1780. For all other generations, the determination of the minimum number of offspring is made in lines 5230 to 5280.

As previously noted, if only one offspring per mating is assumed, the number of offspring produced each generation will be one-half the number of parents. Thus, if the initial parent population consisted of 16 males and 16 females, there would be 8 offspring after the first generation, 4 after the second, 2 after the third and only 1 after the fourth. In order to avoid the vanishing of the population and also avoid using the 'ploy' of assuming that identical twins of opposite sex are born of each mating, it will be assumed that an "average" of two offspring will result from each mating. The student should recall that the previous program describing the genetic evolution of an infinite population implicitly made use of the 'ploy' that identical twins of opposite sex were born of each mating. For small populations, this is not a realistic assumption and it will not be used in this development. Instead, a variable number of offspring will be permitted to be born from a single mating and, on the average over several matings, two offspring per mating will result. In this way, the parent population should remain approximately constant in size. The sex of the offspring will be chosen randomly and siblings will be assumed genetically identical.

The method of selecting the number of offspring, S, resulting from a single mating will be based on the observation that, since two offspring per mating is to be the desired average, the probability of two offspring actually being produced from a single mating should be at least as great as the probability that there result any other number of offspring from a single mating. Also, the probability that no offspring result from a mating should be lower than the probability that two offspring occur from a mating. Finally, the probability that a very large number of offspring result from a mating should be very much lower than the probability that two offspring result from a mating. It will be assumed that no more than seven offspring can result from any single mating. One set of probabilities that accomplish such a distribution of offspring per mating is shown in Table 2 below.

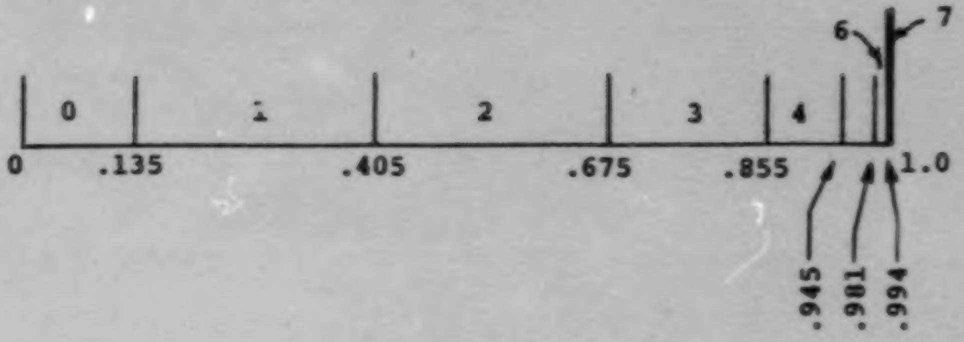
No. of Offspring	Probability of that no. of offspring resulting from a mating
0	0.135
1	0.270
2	0.270
3	0.180
4	0.090
5	0.036
6	0.013
7	0.006

Table 2

Some feel or intuition about the relation of this particular set of probabilities to our problem can be obtained by considering the total number of offspring produced by a very large number of matings for which the number of offspring per mating "behaves" in accord with the listed probabilities. Let the large number of matings be 1000. Then, the fact that 0.135 of the matings will result in no offspring will be interpreted to mean that, of the 1000 matings, 0.135 x 1000, or 135 matings will produce no offspring. Similarly, of the 1000 matings, 0.270 x 1000, or 270, of them will each produce one offspring. Another 270 matings will each produce 2 offspring for a total of 540 offspring. Continuing in this way, it is seen that, 180 matings will each produce 3 offspring for a total of 540 offspring, 90 matings will each produce 4 offspring for a total of 360 offspring, 36 matings will each produce 5 offspring for a total of 180 offspring, 13 matings will each produce 6 offspring for a total of 78 offspring and finally 6 matings will each produce 7 offspring for a total of 42 offspring. The grand total of the number of offspring produced by the 1000 matings is 2000, or an average of 2 offspring per mating. The student should also note that the sum of the probabilities is one and, therefore, one of the numbers 0, 1, 2, 3, 4, 5, 6, or 7, will always be selected as the number of offspring from any mating. There are many sets of probabilities which could yield, "on the average", two offspring per mating. The student can experiment with different sets of probabilities by first setting a limit on the maximum number of offspring to be produced from a mating and then assigning a probability to each of the distinct number of offspring resulting from a single mating. The set of probabilities so assigned must add up to one to insure that all matings are accounted for. The sets of probabilities can then be compared by carrying out a calculation similar to that used above for an assumed 1000 matings. A comparison of the different respective numbers of offspring, as well as the total number of offspring, will provide further insight about the probability distribution.

The set of probabilities shown in Table 2 corresponds to a discrete set of probabilities corresponding to a Poisson distribution whose mean is 2. The student who is familiar with statistics will recognize this. Enough discussion has been presented about the listed set of probabilities so that the student should have some insight about them. Since it is the purpose of this work to minimize the use of formal mathematics, the relation of this topic to classical mathematical statistics will not be further explored here. The interested student is urged to carry out the exploration on his or her own.

In order to mimic the random selection of the number of offspring per mating from such a distribution of probabilities, the technique described in a previous section, pp. 7.32 will be used. This technique is based upon the fact that the random number generator produces a number between 0 and 1. Thus, if the interval (0,1) is partitioned into intervals whose lengths are numerically equal to the respective probabilities, and a random number is selected by the random number generator, this number will "fall" in one of the intervals. The interval containing the random number specifies the number of offspring. The interval (0,1) is completely covered by such a partition because the sum of the probabilities is one. The number corresponding to the end points of such a partition are most easily obtained by forming the cumulative sums of the probabilities. These sums are: 0.135, 0.405, 0.675, 0.855, 0.945, 0.981, 0.994 and 1.000. If these numbers are layed out on the interval (0,1), the resultant intervals will have the desired lengths. This is shown in figure 12.



Partitioning of Unit Interval
for Selecting Number of Offspring

Figure 12

To illustrate the method of selecting a random number of offspring from such a distribution of probabilities, the following numerical example is presented. Suppose RND = 0.963. Since this number is greater than 0.945 and less than 0.981, it lies in the interval corresponding to 5 offspring. Consequently, 5 offspring will be said to result from the mating. The larger intervals correspond to the smaller numbers of offspring and hence will be "landed in" more frequently by a number produced by the RND. This is as it should be because large numbers of offspring are to occur far less frequently than smaller numbers of offspring.

Program Description

Figure 13 contains a list of the notations used in the program and figure 14 is a representation of the overall program organization. The numbers appearing at the upper left of each box in the flowchart refer to the corresponding line, or lines, in the program. Figure 15 is a flowchart of the method for determining the number of offspring resulting from each mating and figure 15a depicts the method of selecting the number of offspring in terms of the program variables.

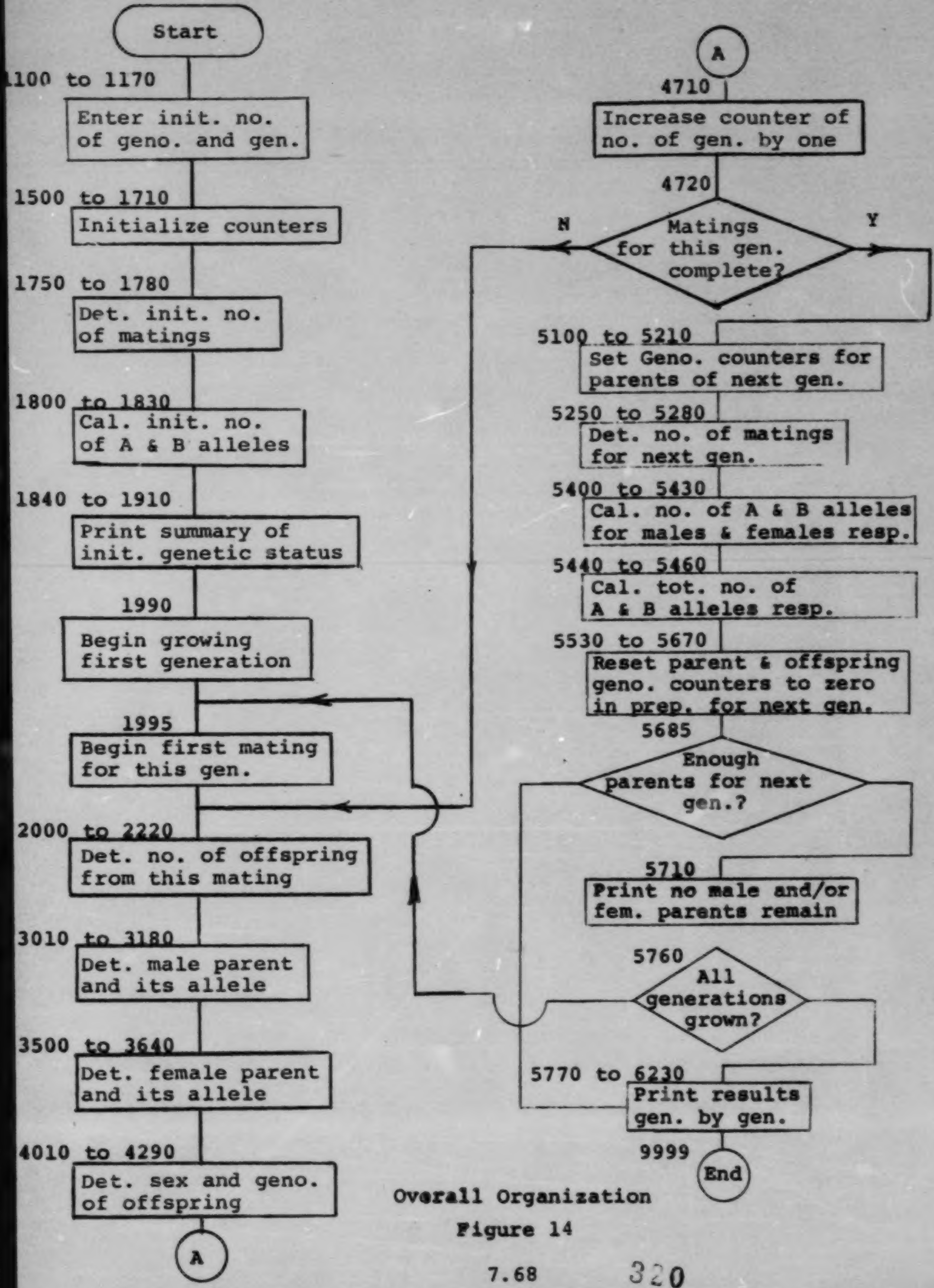
Figure 16 illustrates the method of determining the genotype of the male parent and its allele. Since an identical method is used to determine the female parent and her allele, no flowchart of the procedure is shown. Figure 16a shows this determination in terms of the program variables. The sex and genotype of the offspring are determined in accord with the procedure displayed in figure 17, and figure 17a illustrates the procedure in terms of the program notation. In the program, it is assumed that all siblings are genetically identical. This assumption can be removed by changing the order of the determination of the sex and the genotype.

Most of the program consists of bookkeeping and of the tallying of results whose values are functions of the numbers generated by the random number generator. In attempting to follow the flowcharts, and the program, it is helpful to recall how the production of an offspring from a single mating is to be mimiced. The student is again reminded that, when attempting to gain familiarity with various parts of the program, it is very helpful to remove the RANDOMIZE statement. This will insure that the same set of random numbers will be generated each time the program is run. Thus, replicable numerical results should appear and the same program branches followed each time the program is run. It is also helpful to insert counters at specified points in the program and "to have the counters print out their value"

PROGRAM NOTATION

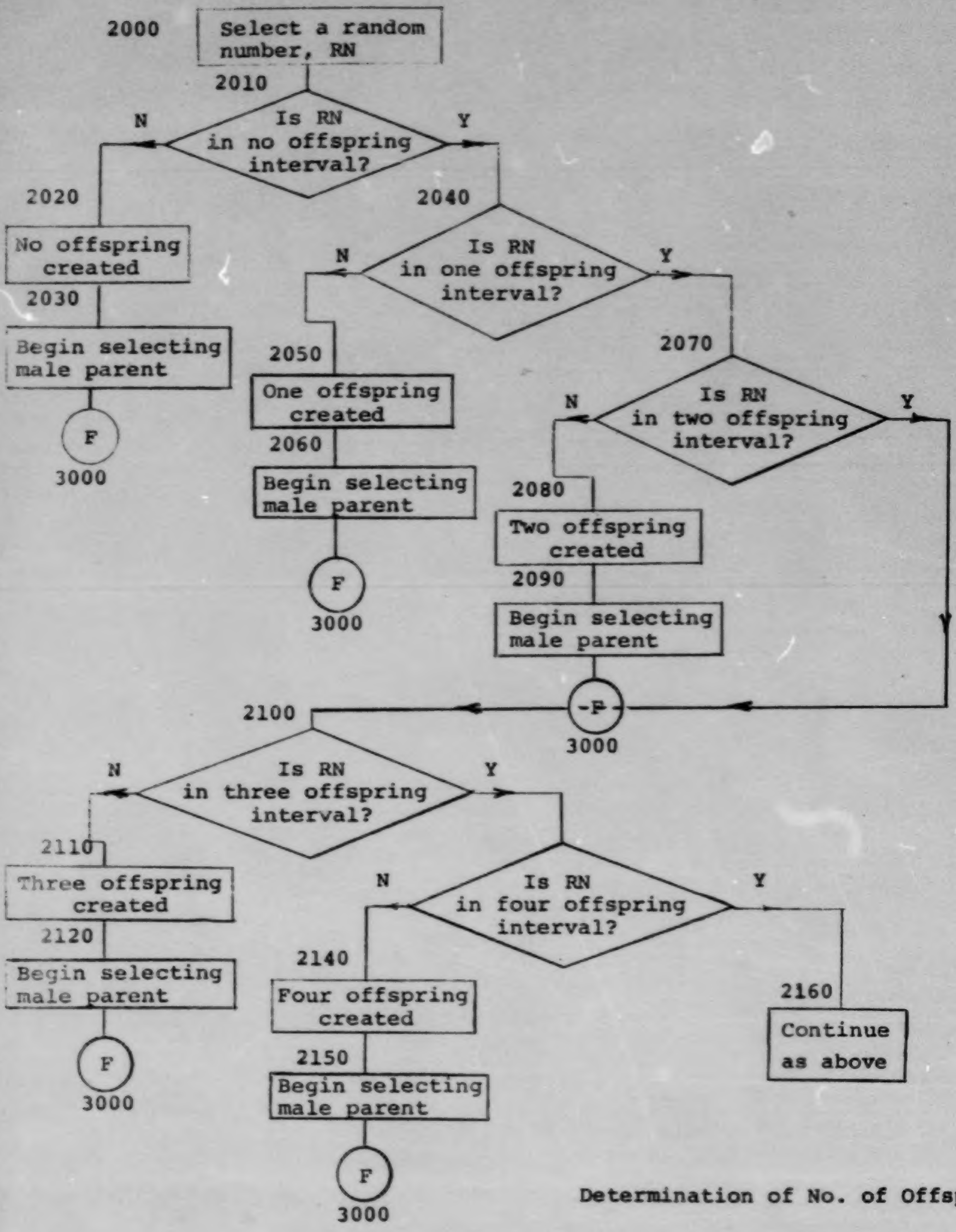
A1(I), A2(I)	No. of A and B male alleles in I^{th} generation
B1(I), B2(I)	No. of A and B female alleles in I^{th} generation
N1(I), N2(I), N3(I)	No. of AA, AB and BB male genotypes in I^{th} generation
N4(I), N5(I), N6(I)	No. of AA, AB and BB female genotypes in I^{th} generation
D1(I), D2(I)	Total no. of A and B alleles in I^{th} generation
D3(I)	Total number of alleles in I^{th} generation
T1(I), T2(I)	Total no. of male and female parents in I^{th} generation
U1(I), U2(I), U3(I)	Counters for no. of AA, AB and BB male parents that have mated in the I^{th} generation
U4(I), U5(I), U6(I)	Counters for no. of AA, AB and BB female parents that have mated in the I^{th} generation
O1(I), O2(I), O3(I)	Counters of male AA, AB and BB offspring genotypes during the I^{th} generation
O4(I), O5(I), O6(I)	Counters of female AA, AB and BB offspring genotypes during the I^{th} generation
C(I)	No. of matings permitted in the I^{th} generation
G	No. of generations
C1	Counter for no. of matings in a generation
S1, S2, ,,, , S7	Cumulative sums for determination of no. of offspring or siblings per mating
M, F	Male and female allele markers
S	No. of offspring or siblings per mating

Figure 13


Start
100 to 1170
Enter init. no. of geno. and gen.
1500 to 1710
Initialize counters
1750 to 1780
Det. init. no. of matings
1800 to 1830
Cal. init. no. of A & B alleles
1840 to 1910
Print summary of init. genetic status
1990
Begin growing first generation
1995
Begin first mating for this gen.
2000 to 2220
Det. no. of offspring from this mating
3010 to 3180
Det. male parent and its allele
3500 to 3640
Det. female parent and its allele
4010 to 4290
Det. sex and geno. of offspring
A
A
4710
Increase counter of no. of gen. by one
4720
N
Matings for this gen. complete?
Y
5100 to 5210
Set Geno. counters for parents of next gen.
5250 to 5280
Det. no. of matings for next gen.
5400 to 5430
Cal. no. of A & B alleles for males & females resp.
5440 to 5460
Cal. tot. no. of A & B alleles resp.
5530 to 5670
Reset parent & offspring geno. counters to zero in prep. for next gen.
5685
Enough parents for next gen.?
5710
Print no male and/or fem. parents remain
5760
All generations grown?
5770 to 6230
Print results gen. by gen.
9999
End

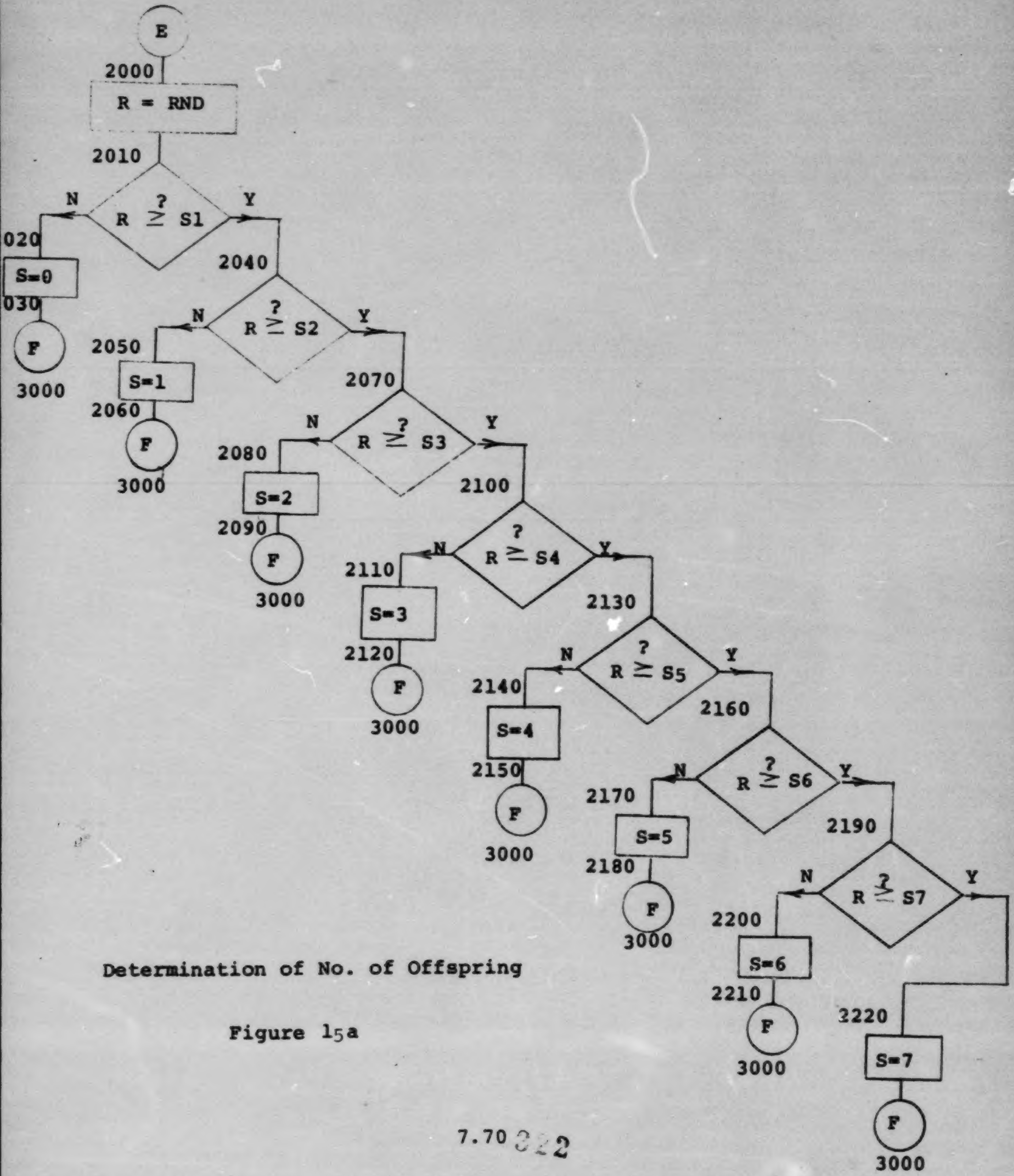
Overall Organization

Figure 14


2000
Select a random number, RN
2010
Is RN in no offspring interval?
N
Y
2020
No offspring created
2030
Begin selecting male parent
F
3000
2040
Is RN in one offspring interval?
N
Y
2050
One offspring created
2060
Begin selecting male parent
F
3000
2070
Is RN in two offspring interval?
N
Y
2080
Two offspring created
2090
Begin selecting male parent
F
3000
2100
Is RN in three offspring interval?
N
Y
2110
Three offspring created
2120
Begin selecting male parent
F
3000
Is RN in four offspring interval?
N
Y
2140
Four offspring created
2150
Begin selecting male parent
F
3000
2160
Continue as above

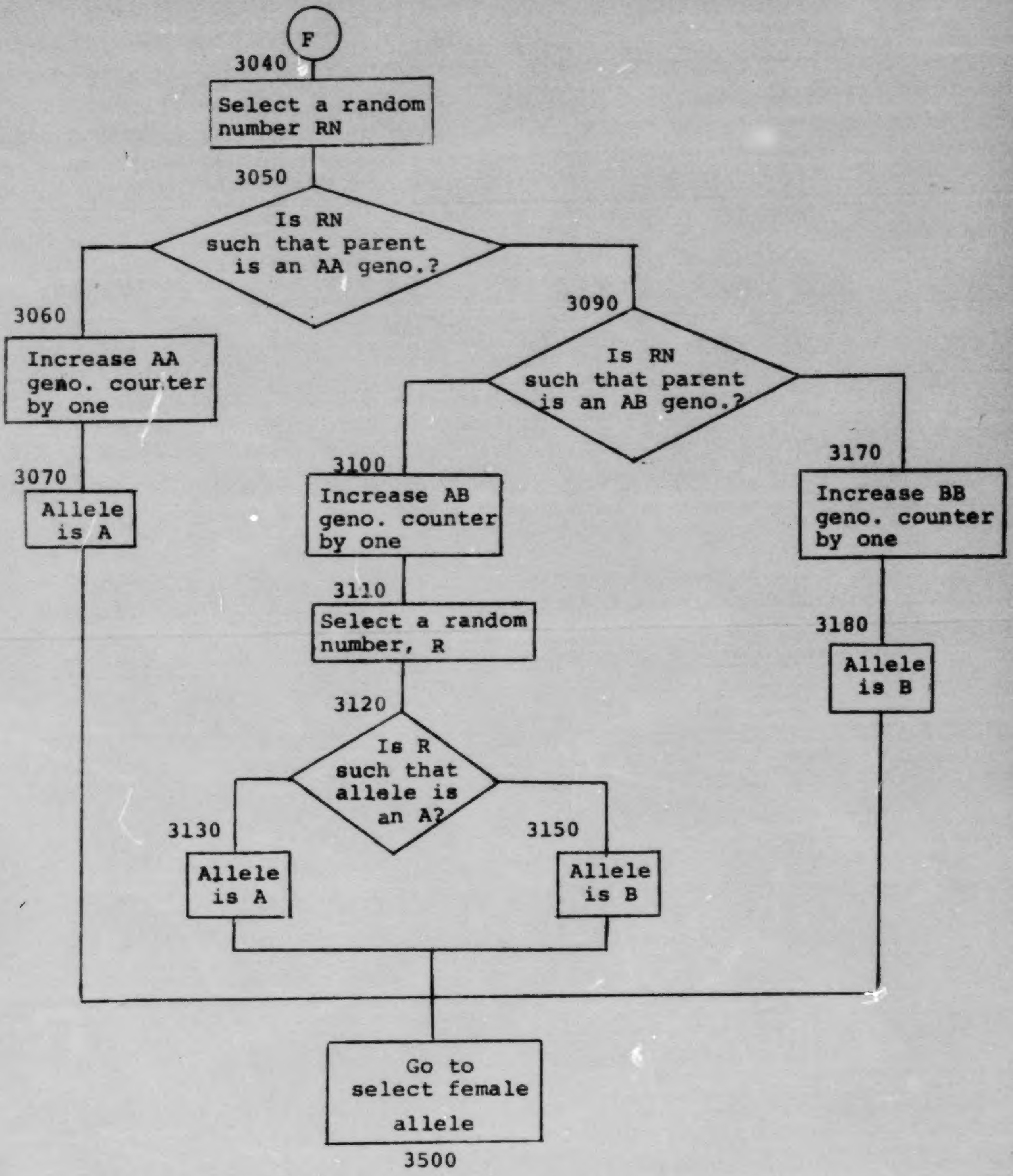
Determination of No. of Offspring

Figure 15


E
2000
R = RND
2010
N
Y
R ≥? S1
020
S=0
030
F
3000
2040
R ≥? S2
2050
S=1
2060
F
3000
2070
R ≥? S3
2080
S=2
2090
F
3000
2100
R ≥? S4
2110
S=3
2120
F
3000
2130
R ≥? S5
2140
S=4
2150
F
3000
2160
R ≥? S6
2170
S=5
2180
F
3000
2190
R ≥? S7
2200
S=6
2210
F
3000
2220
S=7
F
3000

Determination of No. of Offspring

Figure 15a

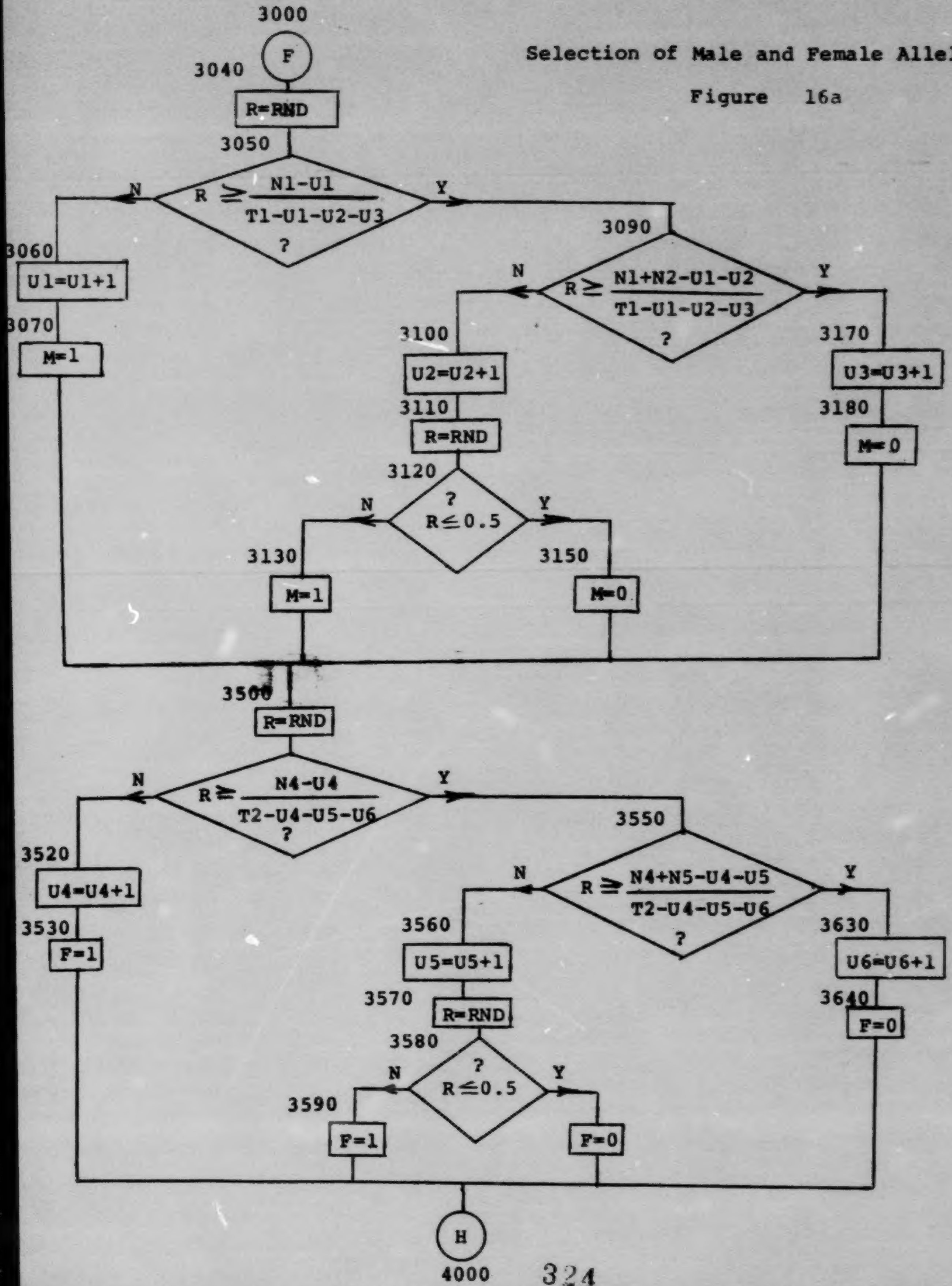

F
3040
Select a random number RN
3050
Is RN such that parent is an AA geno.?
3060
Increase AA geno. counter by one
3070
Allele is A
3090
Is RN such that parent is an AB geno.?
3100
Increase AB geno. counter by one
3110
Select a random number, R
3120
Is R such that allele is an A?
3130
Allele is A
3150
Allele is B
3170
Increase BB geno. counter by one
3180
Allele is B
Go to select female allele
3500

Selection of Male Parent Allele

Figure 16

Selection of Male and Female Alleles

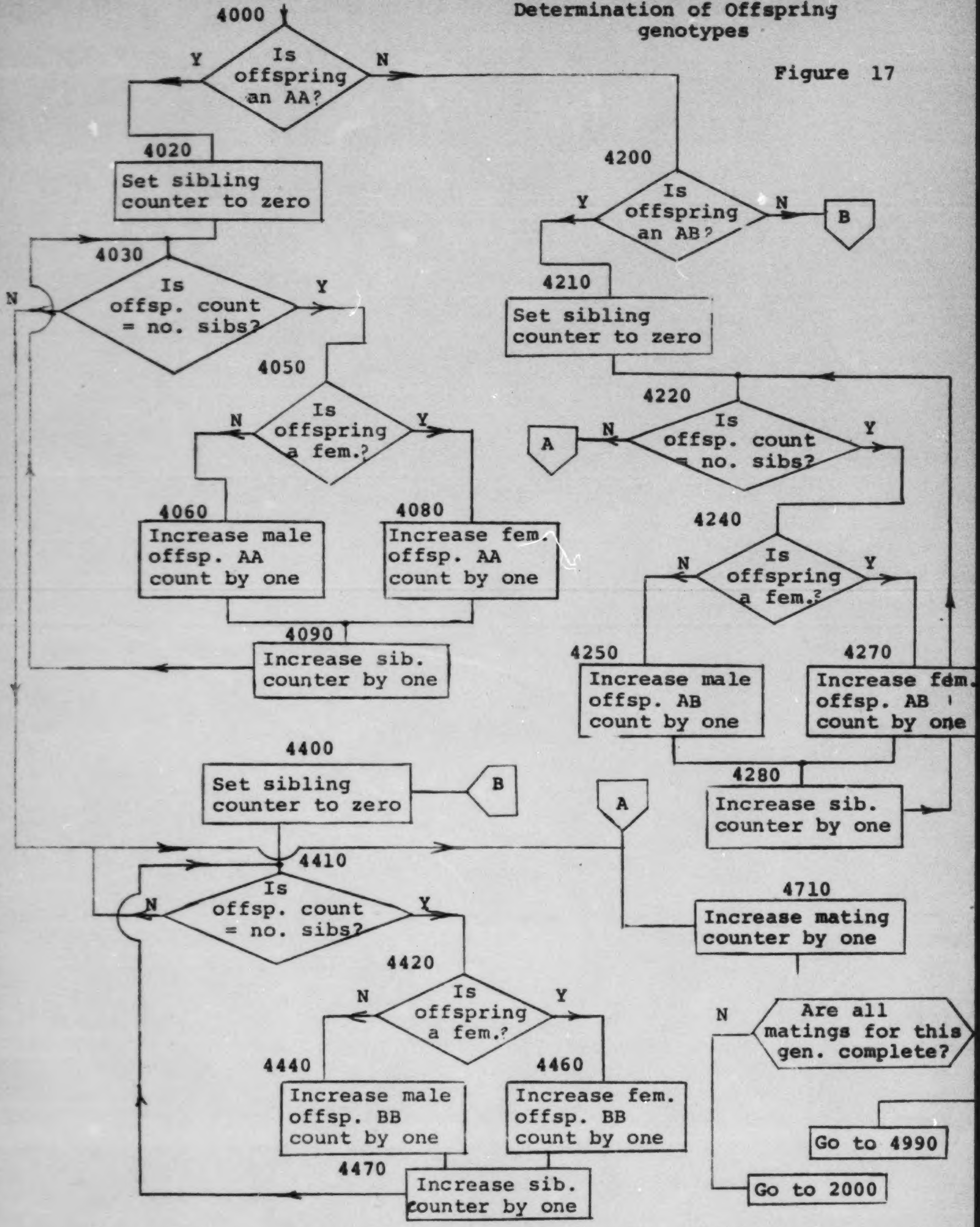
Figure 16a

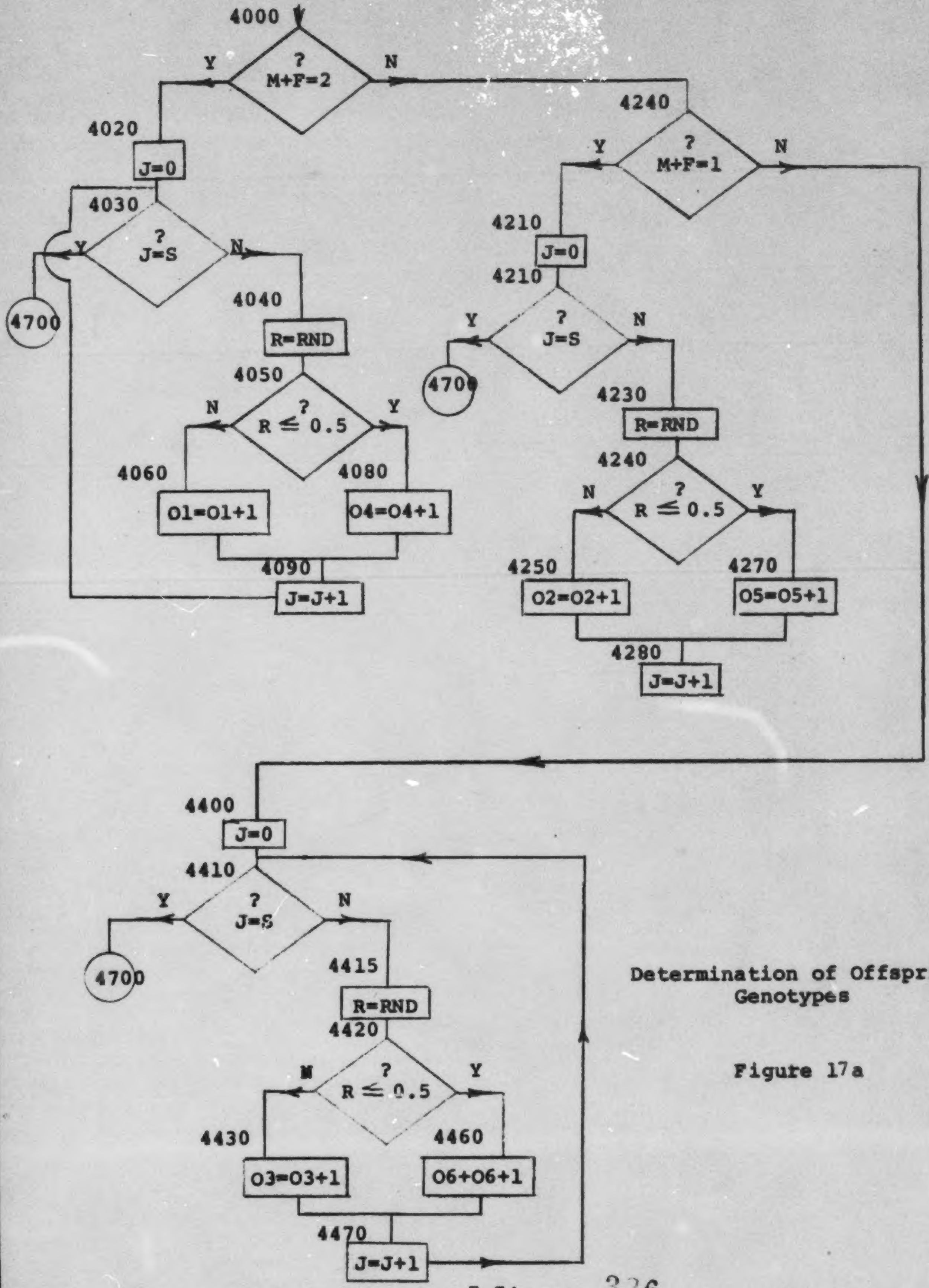

3000
F
3040
R=RND
3050
R ≥ (N1-U1)/(T1-U1-U2-U3) ?
N
Y
3060
U1=U1+1
3070
M=1
3090
R ≥ (N1+N2-U1-U2)/(T1-U1-U2-U3) ?
N
Y
3100
U2=U2+1
3110
R=RND
3120
? R≤0.5
N
Y
3130
M=1
3150
M=0
3170
U3=U3+1
3180
M=0
3500
R=RND
R ≥ (N4-U4)/(T2-U4-U5-U6) ?
N
Y
3520
U4=U4+1
3530
F=1
3550
R ≥ (N4+N5-U4-U5)/(T2-U4-U5-U6) ?
N
Y
3560
U5=U5+1
3570
R=RND
3580
? R≤0.5
N
Y
3590
F=1
F=0
3630
U6=U6+1
3640
F=0
H
4000

7.72

Determination of Offspring genotypes

Figure 17


4000
Is offspring an AA?
Y
N
4020
Set sibling counter to zero
4030
Is offsp. count = no. sibs?
N
Y
4050
Is offspring a fem.?
N
Y
4060
Increase male offsp. AA count by one
4080
Increase fem. offsp. AA count by one
4090
Increase sib. counter by one
4200
Is offspring an AB?
Y
N
B
4210
Set sibling counter to zero
4220
Is offsp. count = no. sibs?
A
N
Y
4240
Is offspring a fem.?
N
Y
4250
Increase male offsp. AB count by one
4270
Increase fem. offsp. AB count by one
4280
Increase sib. counter by one
4400
Set sibling counter to zero
B
A
4410
Is offsp. count = no. sibs?
N
Y
4420
Is offspring a fem.?
N
Y
4440
Increase male offsp. BB count by one
4460
Increase fem. offsp. BB count by one
4470
Increase sib. counter by one
4710
Increase mating counter by one
N
Are all matings for this gen. complete?
Go to 4990
Go to 2000


4000
? M+F=2
Y
N
4020
J=0
4030
? J=S
Y
N
4700
4040
R=RND
4050
? R ≤ 0.5
N
Y
4060
O1=O1+1
4080
O4=O4+1
4090
J=J+1
4240
? M+F=1
Y
N
4210
J=0
4210
? J=S
Y
N
4700
4230
R=RND
4240
? R ≤ 0.5
N
Y
4250
O2=O2+1
4270
O5=O5+1
4280
J=J+1
4400
J=0
4410
? J=S
Y
N
4700
4415
R=RND
4420
? R ≤ 0.5
N
Y
4430
O3=O3+1
4460
O6+O6+1
4470
J=J+1

Determination of Offspring Genotypes

Figure 17a

whenever it is desired. After the program is debugged and tested, these counters and their associated PRINT statements can be deleted from the program.

The following paragraphs contain comments concerning various portions of the program. The paragraphs should be read in conjunction with the program, figure 18, since they refer to the program by line number. The purpose of the comments is to further clarify the program and its structure.

The calculation of the number of male and female A and B alleles appears in lines 1800 to 1830 and in lines 5400 to 5430. These statements express the fact that the homozygote parent alleles, AA and BB, each contribute 2 of the same alleles respectively to the parent allele pool whereas the heterozygote AB parent contributes only one A and one B allele to the parent pool. Lines 1700 and 1710 state that the number of male or female parents is just the sum of the respective AA, AB and BB male or female parent genotypes. Lines 1500 to 1550 initialize the counters which record the number of male or female AA, AB or BB genotypes that have mated as the matings take place during a generation. The operation of initializing a variable sets the variable equal to a desired starting value. In this instance, the desired starting value is zero. The initializing of the counters U1, U2, ,,, U6 insures that there will be no more matings than there are respective genotypes. These counters must be initialized at the start of the growth of each generation. See lines 5550 to 5600.

Lines 1600 to 1650 initialize the counters of the male and female offspring AA, AB and BB genotypes before the start of the growth of the first generation. These counters must also be reset to zero at the start of the growth of each succeeding generation. This is accomplished in lines 5620 to 5670. Because the number of offspring or siblings per mating is not constant, it is possible to have more or less offspring than there are matings or parents. Thus, the respective numbers of offspring genotypes will not usually be equal to the respective numbers of parent genotypes.

```
5 REM          SMALL POPULATION, POPULATION GENETICS PROGRAM
7 REM
20 RANDOMIZE
200 DIM A1(50),A2(50),B1(50),B2(50)
210 DIM N1(50),N2(50),N3(50),N4(50),N5(50),N6(50)
220 DIM T1(50),T2(50),C(50)
230 DIM D1(50),D2(50),D3(50)
400 PRINT "              SMALL POPULATION GENETICS"
401 PRINT
402 PRINT
403 PRINT
500 DATA .135,.405,.675,.855,.945,.981,.994
510 READ S1,S2,S3,S4,S5,S6,S7
1100 PRINT "TYPE INIT. NO. OF  AA,  AB  AND  BB  MALE GENOTYPES"
1110 INPUT N1(0),N2(0),N3(0)
1120 PRINT
1130 PRINT "TYPE THE INIT. NO. OF  AA,  AB  AND  BB  FEMALE GENOTYPES"
1140 INPUT N4(0),N5(0),N6(0)
1150 PRINT
1160 PRINT "TYPE THE NO. OF GENERATIONS,  G"
1170 INPUT G
1180 PRINT
1210 PRINT "      INITIAL GENETIC STATUS OF POPULATION"
1220 PRINT
1230 PRINT
1300 REM     A1(I) & A2(I) ARE NO. OF MALE & FEM.  A ALLELES RESP.
1320 REM
1330 REM     B1(I) & B2(I) ARE NO. OF MALE & FEM.  B ALLELES RESP
1340 REM
1470 REM
1480 REM      LINES 1500 TO 1710 SET INITIAL COUNTERS
1490 REM
1500 LET U1=0
1510 LET U2=0
1520 LET U3=0
1530 LET U4=0
1540 LET U5=0
1550 LET U6=0
1560 REM
1600 LET O1=0
1610 LET O2=0
1620 LET O3=0
1630 LET O4=0
1640 LET O5=0
1650 LET O6=0
```

Figure 18

```
1700 LET T1(0)=N1(0)+N2(0)+N3(0)
1710 LET T2(0)=N4(0)+N5(0)+N6(0)
1720 REM
1730 REM   LINES 1760 TO 1780 CALC. NO. OF MATINGS FOR INIT. PARENTS
1740 REM
1750 IF T1(0)<=T2(0)GO TO 1780
1760 LET C(0)=T2(0)
1770 GO TO 1800
1780 LET C(0)=T1(0)
1800 LET A1(0)=2*N1(0)+N2(0)
1810 LET A2(0)=2*N3(0)+N2(0)
1820 LET B1(0)=2*N4(0)+N5(0)
1830 LET B2(0)=2*N6(0)+N5(0)
1832 LET D1(0)=A1(0)+A2(0)
1834 LET D2(0)=A2(0)+B2(0)
1836 LET D3(0)=D1(0)+D2(0)
1840 PRINT "INIT. NO. OF MALE  A  ALLELES IS";A1(0)
1850 PRINT "INIT. NO. OF FEM.  A  ALLELES IS";B1(0)
1855 PRINT
1860 PRINT "INIT. NO. OF MALE  B  ALLELES IS";A2(0)
1870 PRINT "INIT. NO. OF FEM.  B  ALLELES IS";B2(0)
1875 PRINT
1880 PRINT "TOT. NO. OF INIT.  A  ALLELES IS";A1(0)+B1(0)
1900 PRINT "TOT. NO. OF INIT.  B  ALLELES IS";A2(0)+B2(0)
1905 PRINT
1910 PRINT "INIT. NOS. OF MALE & FEM. PARENTS ARE";T1(0);"  &  ";T2(0)
1915 PRINT
1920 PRINT
1930 REM
1940 REM     LINES 1990 TO 5500 GROW OFFSPRING FOR ONE GENERATION
1959 REM
1960 REM
1970 REM       LINES 2000 TO 2230 SELECT NO. OFFSPRING OF THE MATING
1980 REM
1990 FOR I=0 TO G
1994 LET C1=0
1995 LET C1=0
2000 LET R=RND
2010 IF R>=S1GO TO 2040
2020 LET S=0
2030 GO TO 3000
2040 IF R>=S2GO TO 2070
2050 LET S=1
2060 GO TO 3000
2070 IF R>=S3GO TO 2100
2080 LET S=2
2090 GO TO 3000
2100 IF R>=S4GO TO 2130
2110 LET S=3
2120 GO TO 3000
2130 IF R>=S5GO T 2160
2140 LET S=4
```

Figure 18 (Cont.)

```
2150 GO TO 3000
2160 IF R>=S6GO TO 2190
2170 LET S=5
2180 GO TO 3000
2190 IF R>=S7GO TO 2220
2200 LET S=6
2210 GO TO 3000
2220 LET S=7
2990 REM
3000 REM      LINES 3040 TO 3180 SELECT MALE PARENT ALLELE
3010 REM
3040 LET R=RND
3050 IF R>=(N1(I)-U1)/(T1(I)-U1-U2-U3)GO TO 3090
3060 LET U1=U1+1
3070 LET M=1
3080 GO TO 3500
3090 IF R>=(N1(I)+N2(I)-U1-U2)/(T1(I)-U1-U2-U3)GO TO 3170
3100 LET U2=U2+1
3110 LET R=RND
3120 IF R<=.5GO TO 3150
3130 LET M=1
3140 GO TO 3500
3150 LET M=0
3160 GO TO 3500
3170 LET U3=U3+1
3180 LET M=0
3290 REM
3300 REM       LINES 3500 TO 3640 SELECT FEMALE PARENT ALLELE
3310 REM
3500 LET R=RND
3510 IF R>=(N4(I)-U4)/(T2(I)-U4-U5-U6)GO TO 3550
3520 LET U4=U4+1
3530 LET F=1
3540 GO TO 4000
3550 IF R>=(N4(I)+N5(I)-U4-U5)/(T2(I)-U4-U5-U6)GO TO 3630
3560 LET U5=U5+1
3570 LET R=RND
3580 IF R<=.5GO TO 3610
3590 LET F=1
3600 GO TO 4000
3610 LET F=0
3620 GO TO 4000
3630 LET U6=U6+1
3640 LET F=0
3650 GO TO 4000
3980 REM
3990 REM         LINES 4010 TO 4290 SELECT GENOTYPE OF OFFSPRING
4000 REM
4010 IF (M+F)<>2GO TO 4200
4020 LET J=0
4030 IF J=SGO TO 4700
4040 LET R=RND
4050 IF R<=.5GO TO 4080
4060 LET O1=O1+1
4070 GO TO 4090
4080 LET O4=O4+1
4090 LET J=J+1
4100 GO TO 4030
4200 IF (M+F)<>1G TO 4400
```

Figure 18 (Cont.)

```
4210 LET J=0
4220 IF J=SGO TO 4700
4230 LET R=RND
4240 IF R<=.5GO TO 4270
4250 LET O2=O2+1
4260 GO TO 4280
4270 LET O5=O5+1
4280 LET J=J+1
4290 GO TO 4220
4400 LET J=0
4410 IF J=SGO TO 4700
4415 LET R=RND
4420 IF R<=.5GO TO 4460
4440 LET O3=O3+1
4450 GO TO 4470
4460 LET O6=O6+1
4470 LET J=J+1
4480 GO TO 4410
4485 REM
4690 REM      LINE 4710 & 4720 COUNT NO. OF MATINGS IN THIS GEN.
4700 REM
4710 LET C1=C1+1
4720 IF C1<C(I)GO TO 2000
4990 REM
5000 REM       LINES 5100 TO 5210 SET THE GENOTYPE COUNTERS FOR THIS GEN.
5010 REM
5100 LET N1(I+1)=O1
5110 LET N2(I+1)=O2
5120 LET N3(I+1)=O3
5130 LET N4(I+1)=O4
5140 LET N5(I+1)=O5
5150 LET N6(I+1)=O6
5160 REM
5200 LET T1(I+1)=N1(I+1)+N2(I+1)+N3(I+1)
5210 LET T2(I+1)=N4(I+1)+N5(I+1)+N6(I+1)
5220 REM
5230 REM        LINES 5230 TO 5280 CALC. NO. OF MATINGS FOR NEXT GEN.
5240 REM
5250 IF T1(I+1)<=T2(I+1)GO TO 5280
5260 LET C(I+1)=T2(I+1)
5270 GO TO 5400
5280 LET C(I+1)=T1(I+1)
5370 REM
5380 REM      LINES 5400 TO 5430 CALCULATE NO. OF A AND B ALLELES
5390 REM
5400 LET A1(I+1)=2*O1+O2
5410 LET A2(I+1)=2*O3+O2
5420 LET B1(I+1)=2*O4+O5
5430 LET B2(I+1)=2*O6+O5
5440 LET D1(I+1)=A1(I+1)+B1(I+1)
5450 LET D2(I+1)=A2(I+1)+B2(I+1)
5460 LET D3(I+1)=D1(I+1)+D2(I+1)
```

Figure 18 (Cont.)

```
5500 REM
5510 REM    LINES 5550 TO 5660 RESET PAR. & OFF. GEN. COUNTERS EACH GEN.
5520 REM
5550 LET U1=0
5560 LET U2=0
5570 LET U3=0
5580 LET U4=0
5590 LET U5=0
5600 LET U6=0
5610 REM
5620 LET O1=0
5630 LET O2=0
5640 LET O3=0
5650 LET O4=0
5660 LET O5=0
5670 LET O6=0
5685 IF C(I+1)<>0GO TO 5760
5690 PRINT
5700 PRINT
5705 PRINT
5710 PRINT "ON GEN. NO. ";I+1;"NO MALE AND/OR FEM PARENTS REMAIN."
5720 PRINT "NO. MALE PARENTS =";T1(I+1),"NO. FEMALE PARENTS =";T2(I+1)
5730 PRINT
5740 PRINT
5750 GO TO 5770
5760 NEXT I
5770 PRINT "                          PROGRAM RESULTS          "
5780 PRINT
5785 PRINT
5790 PRINT
5800 PRINT "GEN. NO.      NO. OF MALE PAR.      NO. OF FEM. PAR."
5805 PRINT
5810 FOR I=0 TO G
5820 PRINT " ";I;"                ";T1(I);"                  ";T2(I)
5830 NEXT I
5840 PRINT
5850 PRINT
5980 REM
5990 REM             LINES 6010 TO 6250 PRINT RESULTS
5995 REM
6000 PRINT "                          A ALLELE                    B ALLELE"
6010 PRINT "GEN. NO.          MALE         FEMALE          MALE       FEMALE"
6020 PRINT
6030 FOR I=0 TO G
6040 PRINT I,A1(I),B1(I),A2(I),B2(I)
6050 NEXT I
6080 PRINT
6090 PRINT
6100 PRINT "GEN. NO.    TOT. A ALL.      TOT. B ALL.                TOT. ALL.
"
6120 PRINT
6130 FOR I=0 TO G
6140 PRINT " ";I,D1(I),D2(I),,D3(I)
6150 NEXT I
6160 PRINT
```

Figure 18 (Cont.)

```
6170 PRINT
6180 PRINT "                MALE GENOTYPES             FEMALE GENOTYPES"
6190 PRINT "GEN. NO.     AA    AB    BB               AA    AB    BB"
6200 PRINT
6220 FOR I=0 TO G
6230 PRINT I,N1(I);"  ";N2(I);"  ";N3(I),,N4(I);"  ",N5(I);"  ";N6(I)
6240 NEXT I
6250 PRINT
6260 PRINT
9999 END
```

Figure 18 (Cont.)

Lines 5100 to 5150 set the number of parent genotypes for the next or (I+1)st generation. This explains the use of the index I+1, rather than the index I, in lines 5200 to 5460. All of these lines are specifying the distribution of genotypes and alleles for the parents of the next generation in order that the entire process of growing the offspring of the next generation can be repeated in the next generation.

The student should recall that, if a parent is a homozygote AA or BB, an A or a B allele respectively will descend to be united in the zygote. However, if the parent is a heterozygote AB, it is equally likely that either an A or a B allele will descend. The choice of which allele is to descend is made with the assistance of the random number generator. This explains the use of the RND statement in lines 3110 to 3150 and lines 3570 to 3610.

Miscellaneous Comments

As the DIM statements indicate, the program was written to run for a maximum of 50 generations. By increasing the range of the DIM statements, the program can be run for a larger number of generations. Of course, the computer storage and computer run time will increase accordingly.

The program was written in BASIC because BASIC is the language used in this work. Since most BASIC compilers are interpretive, each statement or line in the program is translated into machine language and then executed by the computer. Thus, if the program is run for many generations, the same lines are translated into machine language again and again for as many times as there are generations to be run. In contrast, a programming language such as FORTRAN, only requires one translation of the FORTRAN language to the machine language. Such a translation is called a compilation. The process of compiling avoids the several translations that are necessary when BASIC is used. In this sense BASIC is inferior to most higher level programming languages such as FORTRAN, APL, COBOL, etc. The program may be more difficult to put together and debug in one of these other higher level languages; however, it will usually run much more efficiently. In fact, if one attempts to combine life table analysis with population genetics, it may well be the case that assembly or machine language may be the most appropriate language to use. However, we reiterate, BASIC was chosen for this work because it is so very easy to learn and is readily available on mini- as well as maxi-computers. Just as the goal of this work is to minimize the use of mathematics, so also is the goal to minimize the use of more sophisticated programming languages and techniques. This is an introductory course which emphasizes conceptualization of quantitative ideas in a single programming language. As the student becomes used to expressing his or her thoughts and ideas in BASIC, and attempts more sophisticated problems, he or she is urged to become familiar with other higher level programming languages and techniques.

Program Results

Figure 19 illustrates the results from a typical run. The initial numbers of male AA, AB and BB genotypes were chosen to be 3, 4, and 5, and the initial numbers of female AA, AB and BB genotypes were selected to be 6, 5, and 4 respectively. The problem was run for 10 generations. As the results indicate, there were no more male parents to mate after 9 generations and consequently no further mating occurred. On other runs with the same starting distribution of genotypes, one or the other of the parent populations would disappear. This would occur after varying numbers of generations. Sometimes the population would continue mating for up to 20 generations whereas on other runs the population would have to discontinue mating after only 5 generations.

The dying out of a parent population after some variable number of generations is due to the random selection of mates and to the assumption of monogamous mating. In particular, the assumption that the number of matings for the subsequent generation is equal to the minimum of the number of the male and the female offspring of the previous generation, serves to accelerate the decrease in the mating population. The entire process is similar to genetic drift. In contrast however, for infinite populations, random mating can induce genetic drift but it cannot result in the disappearance of the population.

It is interesting to run the program for several generations assuming a large mating population. Your author ran the program with an initial male and female genotypic distribution of 100 AA, 200 AB and 100 BB genotypes. A comparison of the number of A alleles with the number of B alleles, generation by generation, revealed that up to the 15th generation, these numbers alternated in being the largest in magnitude. However, after the 15th generation, even though both numbers decreased, the total number of B alleles decreased much faster than did the total number of A alleles. Thus, once the random distribution of

```
GENE2

          SMALL POPULATION GENETICS


TYPE INIT. NO. OF  AA,  AB  AND  BB  MALE GENOTYPES
?3,4,5

TYPE THE INIT. NO. OF  AA,  AB  AND  BB  FEMALE GENOTYPES
?6,5,4

TYPE THE NO. OF GENERATIONS,  G
?10

      INITIAL GENETIC STATUS OF POPULATION


INIT. NO. OF MALE  A  ALLELES IS 10
INIT. NO. OF FEM.  A  ALLELES IS 17

INIT. NO. OF MALE  B  ALLELES IS 14
INIT. NO. OF FEM.  B  ALLELES IS 13

T0T. NO. OF INIT.  A  ALLELES IS 27
TOT. NO. OF INIT.  B  ALLELES IS 27

INIT. NOS. OF MALE & FEM. PARENTS ARE 12   &   15




ON GEN. NO. 9 EITHER NO MALE AND/OR FEM PARENTS REMAIN
NO. MALE PARENTS = 0        NO. FEMALE PARENTS = 0


                PROGRAM RESULTS


GEN.NO.      NO. OF MALE PAR.      NO. OF FEM. PAR.

  0                15                   11
  1                13                   13
  2                15                    5
  3                2                    6
  4                3                    6
  5                5                    2
  6                1                    1
  7                1                    2
  8                0                    0
  9                0                    0
  10                0                    0
```

Figure 19

```
                       A ALLELE                            B ALLELE
GEN.  NO.          MALE         FEMALE            MALE        FEMALE

 0                 10            17               14            13
 1                 13            6                17            16
 2                 8             6                18            18
 3                 8             5                22            5
 4                 2             4                2             8
 5                 6             12               0             0
 6                 10            4                0             0
 7                 2             2                0             0
 8                 2             4                0             0
 9                 0             0                0             0
 10                0             0                0             0


GEN.  NO.     TOT.  A ALL.      TOT.  B ALL.                TOT.  ALL.

  0                24            27                            51
  1                19            33                            52
  2                16            36                            52
  3                13            27                            40
  4                6             10                            16
  5                18            0                             18
  6                14            0                             14
  7                4             0                             4
  8                6             0                             6
  9                0             0                             0
  10               0             0                             0


                   MALE  GENOTYPES                   FEMALE  GENOTYPES
GEN.  NO.        AA     AB     BB                  AA     AB     BB

 0                 3     4      5                  6      5      4
 1                 3     7      5                  1      4      6
 2                 3     2      8                  1      6      6
 3                 2     4      9                  1      3      1
 4                 0     2      0                  0      4      2
 5                 3     0      0                  6      0      0
 6                 5     0      0                  2      0      0
 7                 1     0      0                  1      0      0
 8                 1     0      0                  2      0      0
 9                 0     0      0                  0      0      0
 10                0     0      0                  0      0      0


READY
```

Figure 19 (Cont.)

offspring was such as to produce substantially more of one type of allele than the other type, it was not possible for the latter to again "catch up" with the former type of allele. Consequently, on the 50th generation there were 91 A alleles and 45 B alleles; a ratio of approximately 2:1.

Of course if this problem were to be rerun, the results might show a complete reversal from the previous results. The possible variability of such results suggests that several runs, each with the same starting conditions, should be made and the results averaged. Once this average is determined, the deviation of a particular run from this average can be obtained. The deviation is the quantity of interest. The computer time required to make several runs in order that a reasonable estimate of the average may be obtained may be extensive.

The program can provide insight into many diverse phenomena concerning small populations. Many different initial distributions of parent genotypes can be examined and the effect on the generation to generation variation of the genotypes noted.

From the preceding discussion, it is evident that statistical procedures play a very important and necessary role in genetics. Such procedures are useful in determining sample sizes, the degree of correlation of results, statistical measures such as the mean, the average, the variance, etc. Statistics also forms the basis of the analysis of the bias or non-randomness of sets of numbers produced by random number generators. For these reasons, those students who are interested in further pursuing the study of genetics are urged to become familiar with the techniques of statistical analysis.

Another Program Alteration

This section considers a modification of the first population genetics program to include selectivity and survivability. Biologists know from experiments that, if the reproductive and the survival capabilities of a given genotype are limited, after several generations that genotype will either disappear from the population or the genotypic ratios of successive generations will change.

Our discussion of reproductive and survival capability will be quite restrictive. It will be assumed that the controlling factors of reproduction and survival are:

(1) the capability of a parent genotype to mate, and

(2) the capability of an offspring zygote, once formed, to survive.

It will further be assumed that a measure of these capabilities can be described with the aid of selection coefficients and survivability coefficients. These coefficients are defined below and their definitions differ from those commonly given in genetics texts. However, these definitions are adopted because they seem quite natural for this discussion and this setting.

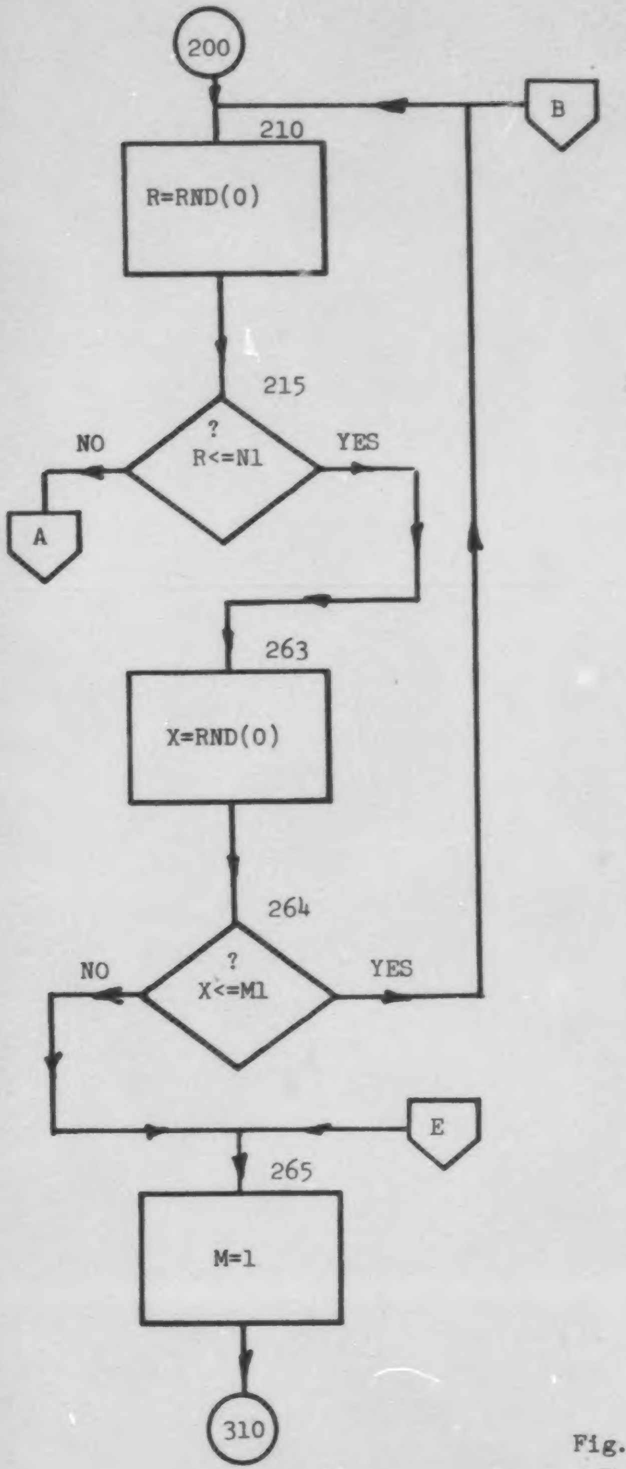
In order to investigate the selection and survival capabilities, the first population genetics program will be altered to include them. To do this, it is convenient to introduce the idea of a selection coefficient. A selection coefficient is a measure of selection against a characteristic or phenomenon, i.e. its action is just the opposite of a preference for something. The selection coefficient is a positive number whose magnitude is less than unity and it is used in the following way. Let M1 denote the effective mating selection coefficient of the male AA genotype, then M1=0.4 means that 0.6, or 60%, of the male AA will effectively mate. Similarly, M1=0 means all male AA genotypes will effectively mate and M1=1 means no male AA genotype mating will result in an offspring. In a similar manner, M2 and M3 are the respective measures against mating effectiveness by the male AB and BB genotypes. Analogously, F1, F2 and F3 will denote the three female selection coefficients. The inclusion of the ability to simulate the survivability of offspring of a prescribed genotype is accomplished by introducing the coefficients C1, C2 and C3. These are called the survivability coefficients for the offspring with AA, AB, and BB genotypes respectively. The survivability coefficient is interpreted in a manner directly opposite to that of the selectivity coefficient. Thus, C2=0.40 means that only 0.40, or 40%, of the AB offspring will survive to become parents for the next generation. Similarly, C2=1 means all AB offspring will survive to become parents for the next generation.

Figures 20a,b,c are flowcharts describing the selectivity of the male parents and figures 21a,b,c are the corresponding flowcharts depicting the selectivity of the female parents. The selectivity of the offspring and the tallying of the number of offspring of the three different genotypes is portrayed in figures 22a and 22b respectively. The program is listed in figures 23a,b. As in the previous flowcharts, the numbers appearing adjacent to the enclosures refer to the corresponding lines of the programs. The "English" language BASIC notation versions of the flowchart have been combined to save space. It is hoped this will not confuse the student.

In the flowcharts, and in the program, the following additional notation has been used:

A1, A2 and A3	denote the no. of AA, AB and BB genotypes in the original population
G	denotes the no. of generations
C	denotes the no. of offspring per generation

Selectivity of Male Parents


200
B
210
R=RND(0)
215
?
R<=N1
NO
YES
A
263
X=RND(0)
264
?
X<=M1
NO
YES
E
265
M=1
310

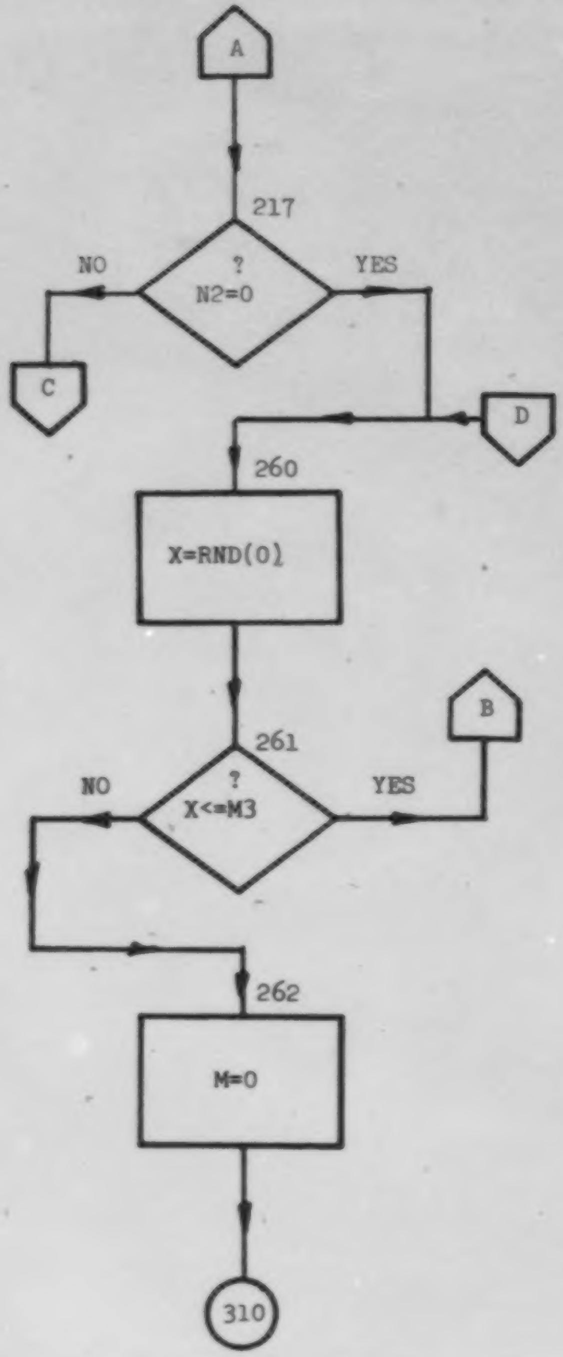
215 Is male parent an AA genotype?

264 Is male AA parent selected against?
(If yes, select another male parent.)

265 Male allele is an A.

Fig. 20a
7.90

Selectivity of Male Parents (cont.)


A
217
NO
?
N2=0
YES
C
D
260
X=RND(0)
B
261
NO
?
X<=M3
YES
262
M=0
310

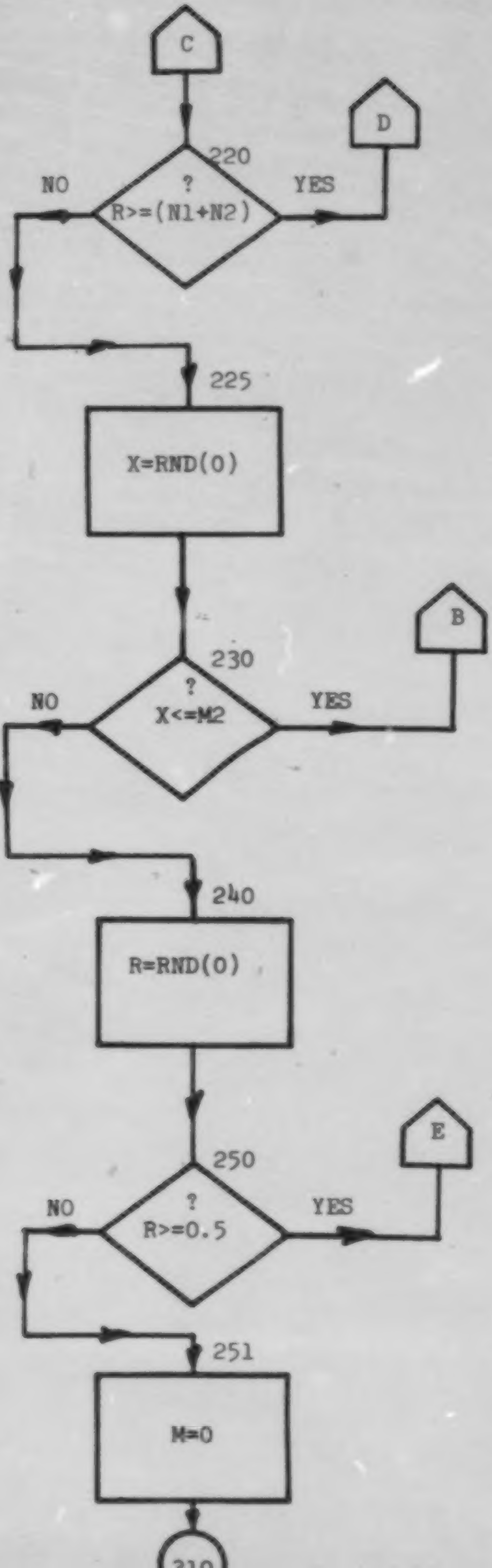
217 Are there no AB male parents? (A yes answer implies that the male parent is a BB genotype.)

261 Is male BB parent selected against? (If yes, select another male parent.)

262 Male allele is a B.

Fig. 20b
7.91

Selectivity of Male Parents (cont.)


C
D
220
NO
?
R>=(N1+N2)
YES
225
X=RND(0)
B
230
NO
?
X<=M2
YES
240
R=RND(0)
E
250
NO
?
R>=0.5
YES
251
M=0
310

220 Is male parent a BB genotype? (A no answer implies that the male parent is an AB genotype; a yes answer implies that the male parent is a BB genotype.)

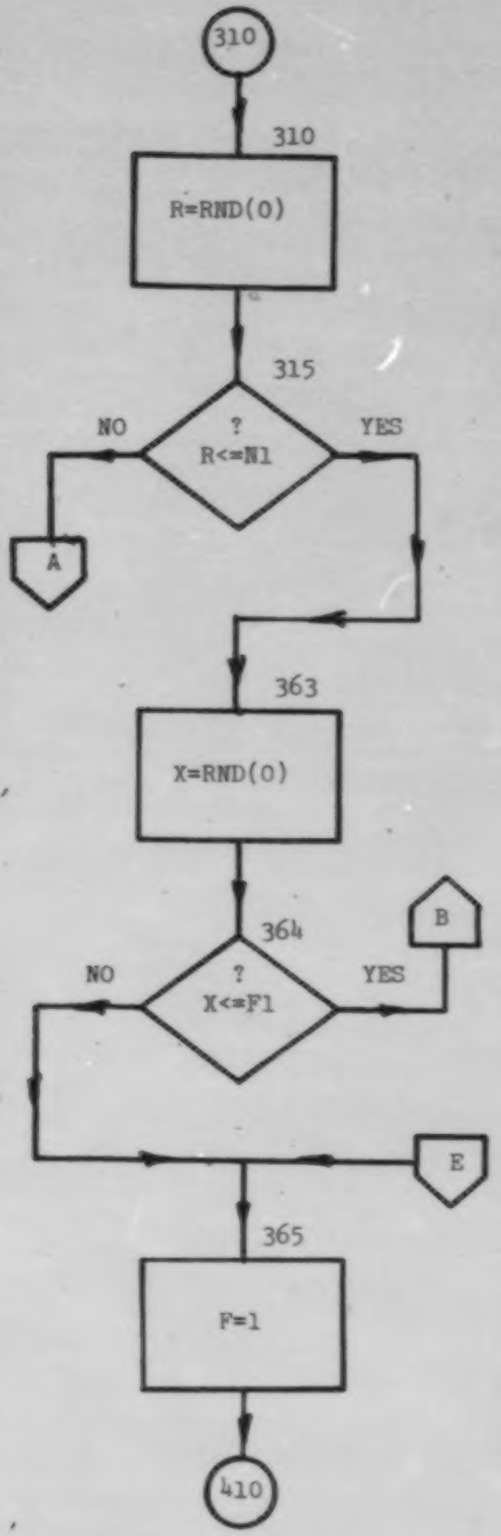
230 Is AB male parent selected against?

250 Is allele of male AB parent an A?

251 Male allele is a B.

Fig. 20c

Selectivity of Female Parents

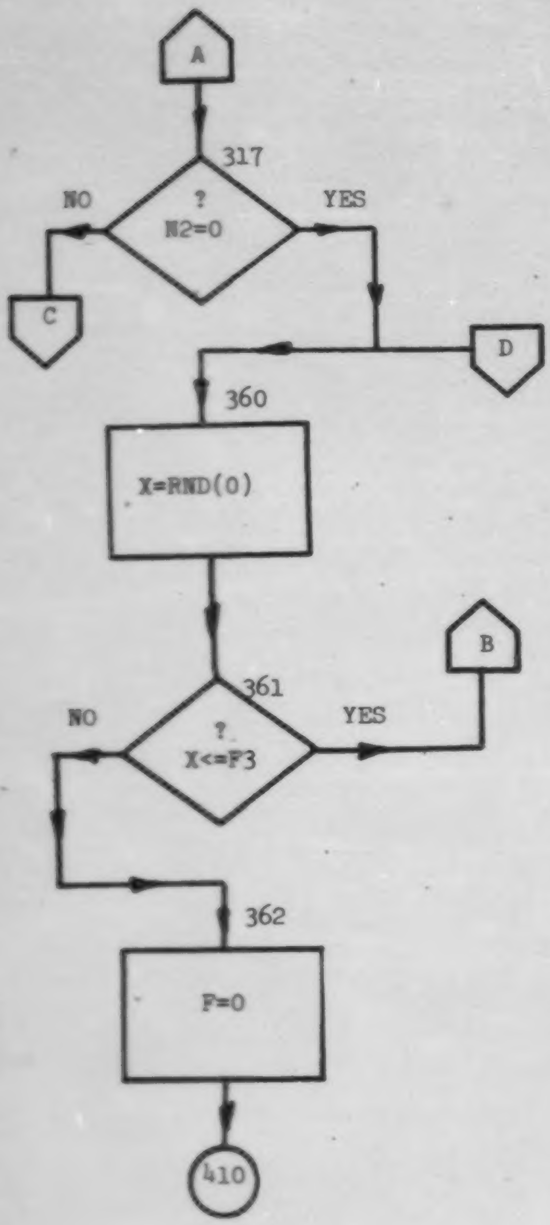

310
310
R=RND(0)
315
NO
?
R<=N1
YES
A
363
X=RND(0)
B
364
NO
?
X<=F1
YES
E
365
F=1
410

364 Is AA female parent selected against? (If yes, start mating process over.)

365 Female allele is A.

Fig. 21a

Selectivity of Female Parents (cont.)


A
317
NO
?
N2=0
YES
C
D
360
X=RND(0)
B
361
NO
?
X<=F3
YES
362
F=0
410

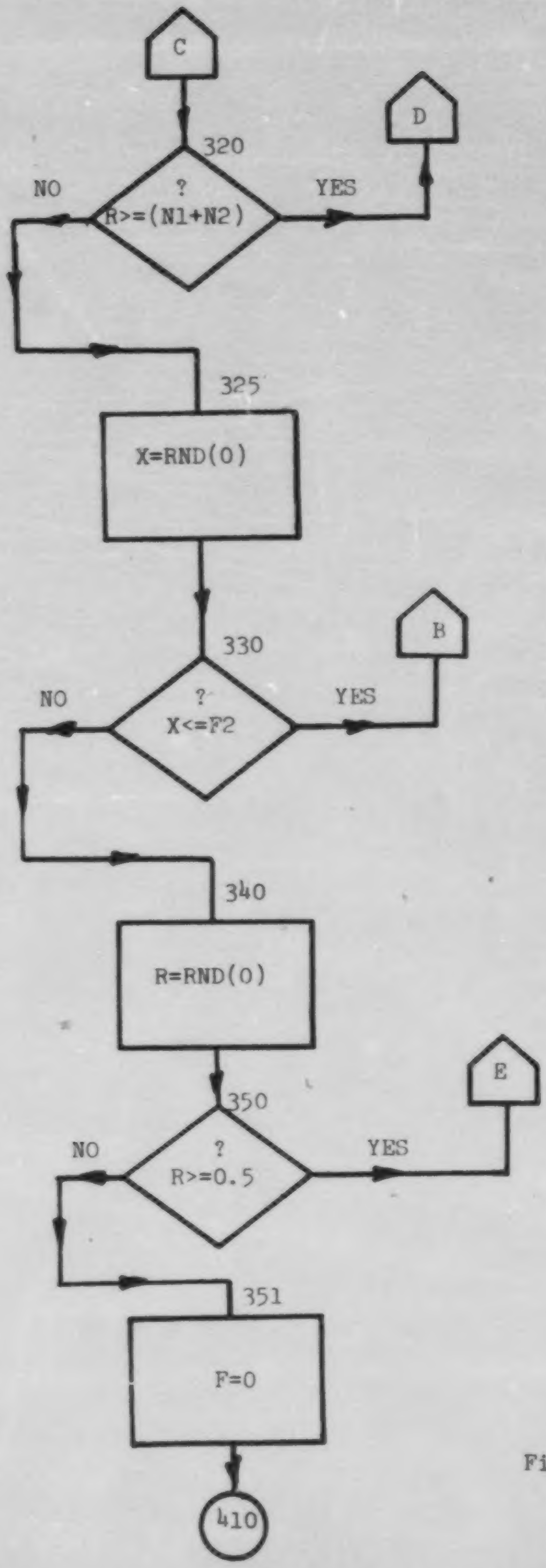
317 Are there no female AB parents? (A yes answer implies female parent is BB.)

361 Is BB female parent selected against? (If yes, start mating process over.)

362 Female allele is B.

Fig. 21b

Selectivity of Female Parents (cont.)


C
D
320
NO
?
R>=(N1+N2)
YES
325
X=RND(0)
330
B
NO
?
X<=F2
YES
340
R=RND(0)
E
350
NO
?
R>=0.5
YES
351
F=0
410

320 Is female parent a BB genotype? (A no answer implies female parent is an AB genotype, a yes answer implies female parent is BB genotype.)

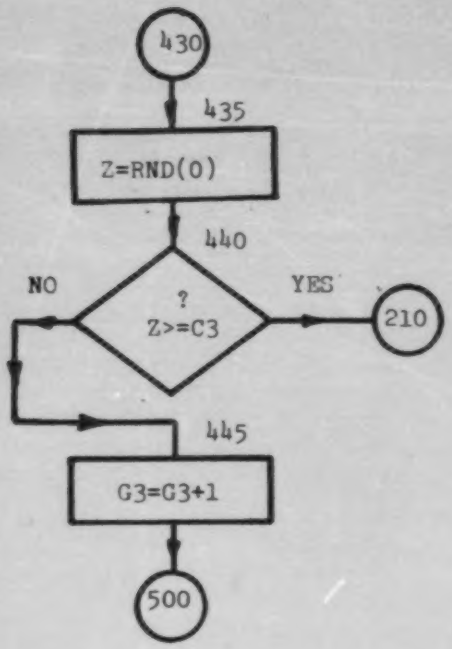
330 Is AB female parent selected against? (If yes, start mating process over.)

350 Is allele of female AB parent an A?

351 Female allele is a B.

Fig. 21c

Survivability of Offspring

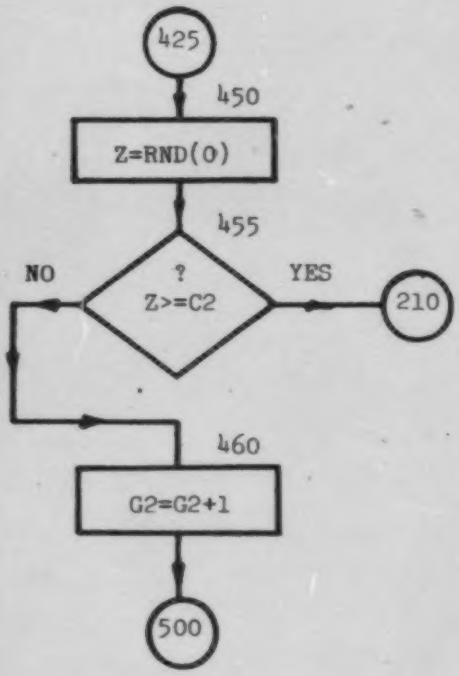

430
435
Z=RND(0)
440
NO
?
Z>=C3
YES
210
445
G3=G3+1
500

430 Offspring is BB genotype.

440 Does BB offspring survive?

210 BB offspring does not survive; go back and "grow" another offspring.

445 One more BB offspring is created.


425
450
Z=RND(0)
455
NO
?
Z>=C2
YES
210
460
G2=G2+1
500

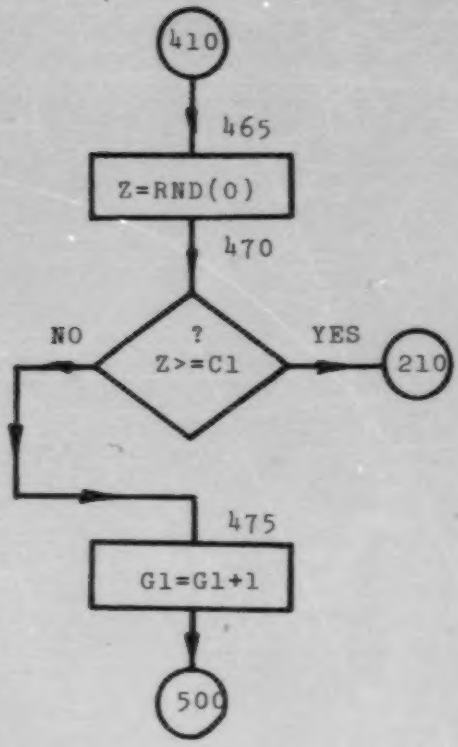
425 Offspring is AB genotype.

455 Does AB offspring survive?

210 AB offspring does not survive; go back and "grow" another offspring.

460 One more AB offspring is created.

Fig. 22a


410
465
Z=RND(0)
470
?
Z>=C1
NO
YES
210
475
G1=G1+1
500

410 Offspring is AA genotype.

470 Does AA offspring survive?

210 AA offspring does not survive;
go back and "grow" another offsprin

475 One more AA offspring is created.

Fig. 22b

RB052

```
1 REM          GENETICS PROGRAM, MULTI-GENERATIONS, SELECTIVITY
3 REM          AND SURVIVABILITY,  INPUT ORIG. NO. GENOTYPES
4 REM
5 REM
6 REM
7 REM          LINES 10-46 PROVIDE FOR INPUT
8 REM
10 PRINT "A1=NO. OF  AA  GENOTYPES IN ORIG.  POP."
11 PRINT "TYPE  A1"\INPUT A1
12 PRINT
15 PRINT "A2=NO. OF  AB  GENOTYPES IN ORIG. POP."
16 PRINT "TYPE A2"\INPUT A2
17 PRINT
20 PRINT "A3=NO.  OF  BB  GENOTYPES IN ORIG. POP."
21 PRINT "TYPE A3"\INPUT A3
22 PRINT
25 PRINT "G=NO.  OF GENERATIONS TO RUN PROGRAM"
26 PRINT "TYPE G"\INPUT G
27 PRINT
30 PRINT "C=NO.  OF OFFSPRING PER GENERATION"
31 PRINT "TYPE C"\INPUT C
32 PRINT
35 PRINT "M1,  M2,  M3  ARE MALE SEL.  COEFFS."
36 PRINT "TYPE M1, M2, M3"\INPUT M1,M2,M3
37 PRINT
40 PRINT "F1,  F2,  F3  ARE FEMALE SEL.  COEFFS."
41 PRINT "TYPE F1, F2, F3"\INPUT F1,F2,F3
42 PRINT
45 PRINT "C1,  C2,  C3  ARE OFFSPRING SURVIVABILITY COEFFS."
46 PRINT "TYPE C1, C2, C3"\INPUT C1,C2,C3
47 PRINT
48 PRINT
106 REM
107 REM M1, M2, M3, F1, F2, F3 ARE THE MALE AND FEM.  SEL.  COEFFS.
108 REM  EXAMPLE:  M1=0.4 MEANS 60% OF  AA  GENOTYPES WILL MATE
109 REM  EXAMPLE:  M1=1.0 MEANS COMPLETE SELECTION AGAINST AN
110 REM               AA MALE PARENT
111 REM
112 REM       C1, C2, C3 ARE OFFSPRING SURVIVABILITY COEFFICIENTS
113 REM       EXAMPLE:  C2=0.45 MEANS 45% OF AB OFFSPRING WILL SURVIVE
114 REM
115 RANDOMIZE
120 LET N=A1+A2+A3
130 LET N1=A1/N\LET N2=A2/N\LET N3=A3/N
131 PRINT "THE INITIAL PROPORTION OF  AA  GENOTYPES IS"N1
132 PRINT
133 PRINT "THE INITIAL PROPORTION  AB  GENOTYPES IS"N2
134 PRINT
135 PRINT "THE INITIAL PROPORTION OF  BB  GENOTYPES IS"N3
136 PRINT
140 LET G1=A1\LET G2=A2\LET G3=A3
141 REM
```

Figure 23a

```
142 REM  G1, G2, G3 ARE THE NO. OF AA, AB, BB GENOTYPES RESPECTIVELY
143 REM
145 FOR I=1 TO G
146 LET N1=G1/N\LET N2=G2/N\LET N3=G3/N
150 LET G1=0\LET G2=0\LET G3=0
170 FOR K=1 TO C
189 REM
190 REM        LINES 210-365 PICK EACH PARENT AND IT'S ALLELE
191 REM
200 REM       LINES 210-265 PICK A MALE PARENT AND HIS ALLELE
201 REM
210 LET R=RND(0)
215 IF R<=N1GO TO 263
217 IF N2=0GO TO 260
220 IF R>=(N1+N2)GO TO 260
225 LET X=RND(0)
230 IF X<=M2GO TO 210
240 LET R=RND(0)
250 IF R>=.5GO TO 265
251 LET M=0\GO TO 310
260 LET X=RND(0)
261 IF X<=M3GO TO 210
262 LET M=0\GO TO 310
263 LET X=RND(0)
264 IF X<=M1GO TO 210
265 LET M=1\GO TO 310
299 REM
300 REM          LINES 310-365 PICK A FEMALE PARENT AND HER ALLELE
301 REM
310 LET R=RND(0)
315 IF R<=N1GO TO 363
317 IF N2=0GO TO 360
320 IF R>=(N1+N2)GO TO 360
325 LET X=RND(0)
330 IF X<=F2GO TO 210
340 LET R=RND(0)
350 IF R>=.5GO TO 365
351 LET F=0\GO TO 410
360 LET X=RND(0)
361 IF X<=F3GO TO 210
362 LET F=0\GO TO 410
363 LET X=RND(0)
364 IF X<=F1GO TO210
365 LET F=1\GO TO 410
399 REM
```

Figure 23b

```
400 REM               LINES 410425 DETERMINE THE GENOTYPE
401 REM
410 IF M+F=2GO TO 465
425 IF M+F=1GO TO 450
429 REM
430 REM         LINES 445, 460 AND 475 COUNT THE GENOTYPES
431 REM
435 LET Z=RND(0)
440 IF Z>=C3GO TO 210
445 LET G3=G3+1\GO TO 500
450 LET Z=RND(0)
455 IF Z>=C2GO TO 210
460 LET G2=G2+1\GO TO 500
465 LET Z=RND(0)
470 IF Z>=C1GO TO 210
475 LET G1=G1+1\GO TO 500
500 NEXT K
501 REM
502 REM        LINES 508-585 PROVIDE FOR OUTPUT
503 REM
506 PRINT
507 PRINT
508 PRINT "THE GENERATION NUMBER IS"I+1
510 PRINT "THE NUMBER OF OFFSPRING IS"C
515 PRINT
520 PRINT "THE NUMBER OF  AA  GENOTYPES IS"G1
525 PRINT
530 PRINT "THE NUMBER OF  AB  GENOTYPES IS"G2
535 PRINT
540 PRINT "THE NUMBER OF  BB  GENOTYPES IS"G3
545 PRINT
550 LET R1=G1/G1\LET R2=G2/G1\LET R3=G3/G1
555 PRINT
565 PRINT "THE RATIO OF  AA  TO  AA  GENOTYPES IS"R1
570 PRINT
575 PRINT "THE RATIO OF  AB  TO  AA  GEONTYPES IS"R2
580 PRINT
585 PRINT "THE RATIO OF  BB  TO  AA  GENOTYPES IS"R3
590 PRINT
593 LET N=G1+G2+G3
595 NEXT I
600 END

READY
```

Figure 23c

N	denotes the sum of the genotypic ratios
N1, N2 and N3	denote the AA, AB and BB genotypic ratios, and
G1, G2, and G3	denote the AA, AB and BB genotypes of the parents for the next generation.

The plan of the computer program is as follows:

From an initial specification of the number of AA, AB or BB parental genotypes (the number is assumed to be the same for both the male and female parents). An AA, AB or BB male parent is selected with the aid of the random number generator. The random number generator is again used with the appropriate selection coefficient M1, M2 or M3 to decide whether or not the male mating is sterile, i.e. whether or not a male gamete descends to form a zygote. (See inst. 225-265). If a male gamete is not forthcoming, a new male parent is selected and the process is repeated. The female gamete is obtained in the same way except that, if a gamete from the female is not forthcoming, the entire process is repeated, i.e. a new male parent is selected and the total process repeated until both a male and female gamete are forthcoming and hence, it is postulated that a zygote is formed. The survivability of the zygote to become a parent for the next generation is simulated by using the random number generator in conjunction with the survivability coefficients C1, C2 and C3. (See insts. 445, 460 and 475 where the counting of the offspring genotypes is also accomplished). If a zygote does not survive, the entire process of growing a new zygote is repeated. As in the previous program, the genotypical description of the offspring is used as the genotypical description of both the male and female parents for the next generation because it is assumed that the offspring are the only parents in the next generation.

The student is again reminded that the following assumptions are made in the program:

(1) The genotypic ratios of the male and female parents are the same and are equal to the corresponding genotypic ratios of the offspring of the previous generation.

(2) The population is infinite. Thus, a large number of offspring must be grown each generation to assure a reasonable estimate of the genotypic ratios for the parents of the next generation.

The construction of the program assumed that:

(1) There is no need for subscripts. Consequently, they are not used. This is in contrast to the previous program. The use of subscripted variables is frequently a matter of taste.

(2) The student will not be confused by the use of multiple statements occurring in a single line of the program. This has been done to restrict the length of the program. Most BASIC compilers have this capability. If the compiler on your computer does not have this capability, the necessary modifications that must be made to the program are evident and should not be difficult to make. Multiple statement lines are indicated wherever the symbol "\" appears. Examples of multiple statement lines are: line 11, line 130, line 250, etc.

The program as written, contains much more interesting information than is printed out. For example, counters could be inserted to determine the number of AA, AB and BB male attempts at mating that had to occur in order to successfully mate with a female. Counters could also be inserted in the program in order to determine the number of attempted male matings necessary to produce a zygote which would survive. In an analogous manner, similar facts could readily be obtained about female attempts at matings.

Many modifications could readily be made to the program to investigate other interesting genetic phenomena. For example, a simple modification of the program to permit the slow change of the selectivity or survivability coefficients each generation could possible simulate the effect of variable mutation. The assumption of equal genotypical descriptions for both the male and female parents could be removed, although for infinite populations this may not be too biologically meaningful. The selection and survivability coefficients could be made sex dependent.

The output from a typical program run is shown in figures 24a and 24b. The initial numbers of AA, AB and BB genotypes were chosen to be 1, 2 and 1 respectively. Since it is the genotypic ratios that determine the selection of mating pairs, it was convenient to enter an equilibrium distribution using small numbers. The 10,000 offspring were grown to provide reasonable accuracy and stability to the results. There is no selection against either an AA or an AB genotype of either sex of parent. However, the initial conditions did specify a selection of 50% against male BB parents and selection of 25% against BB female parents. Thus, there is distinct pressure against the successful mating of a BB genotype, male or female. The BB offspring is subject to a 75% survivability whereas the remaining genotypes are assumed to all survive. In this run, it is evident that the entire mating process is such as to discriminate against the BB genotype.

This particular run illustrates the use of just one possible set of selection and survivability coefficients. Other sets of coefficients could have been used and would have indicated different degrees of pressure on the same or other genotypes. Your

```
RB052

A1=NO.  OF  AA  GENOTYPES IN ORIG.  POP.
TYPE  A1
?1

A2=NO. OF  AB  GENOTYPES IN ORIG. POP.
TYPE A2
?2

A3=NO.  OF  BB  GENOTYPES IN ORIG. POP.
TYPE A3
?1

G=NO.  OF GENERATIONS TO RUN PROGRAM
TYPE G
?4

C=NO.  OF OFFSPRING PER GENERATION
TYPE C
?10000

M1,  M2,  M3  ARE MALE SEL.  COEFFS.
TYPE M1, M2, M3
?0, 0, .5

F1,  F2,  F3  ARE FEMALE SEL.  COEFFS.
TYPE F1, F2, F3
?0, 0, .25

C1,  C2,  C3  ARE OFFSPRING SURVIVABILITY COEFFS.
TYPE C1, C2, C3
?1, 1, .8


THE INITIAL PROPORTION OF  AA  GENOTYPES IS .25

THE INITIAL PROPORTION  AB  GENOTYPES IS .5

THE INITIAL PROPORTION OF  BB  GENOTYPES IS .25


THE GENERATION NUMBER IS 2
THE NUMBER OF OFFSPRING IS 10000

THE NUMBER OF  AA  GENOTYPES IS 3065

THE NUMBER OF  AB  GENOTYPES IS 5241

THE NUMBER OF  BB  GENOTYPES IS 1694

THE RATIO OF  AA  TO  AA  GENOTYPES IS 1

THE RATIO OF  AB  TO  AA  GEONTYPES IS 1.70995

THE RATIO OF  BB  TO  AA  GENOTYPES IS .552692
```

Output

Figure 24a

Output (Cont.)

```
THE GENERATION NUMBER IS 3
THE NUMBER OF OFFSPRING IS 10000

THE NUMBER OF  AA  GENOTYPES IS 3793

THE NUMBER OF  AB  GENOTYPES IS 4925

THE NUMBER OF  BB  GENOTYPES IS 1282

THE RATIO OF  AA  TO  AA  GENOTYPES IS 1

THE RATIO OF  AB  TO  AA  GEONTYPES IS 1.29844

THE RATIO OF  BB  TO  AA  GENOTYPES IS .337991


THE GENERATION NUMBER IS 4
THE NUMBER OF OFFSPRING IS 10000

THE NUMBER OF  AA  GENOTYPES IS 4499

THE NUMBER OF  AB  GENOTYPES IS 4582

THE NUMBER OF  BB  GENOTYPES IS 919

THE RATIO OF  AA  TO  AA  GENOTYPES IS 1

THE RATIO OF  AB  TO  AA  GEONTYPES IS 1.01845

THE RATIO OF  BB  TO  AA  GENOTYPES IS .204268


THE GENERATION NUMBER IS 5
THE NUMBER OF OFFSPRING IS 10000

THE NUMBER OF  AA  GENOTYPES IS 5031

THE NUMBER OF  AB  GENOTYPES IS 4210

THE NUMBER OF  BB  GENOTYPES IS 759

THE RATIO OF  AA  TO  AA  GENOTYPES IS 1

THE RATIO OF  AB  TO  AA  GEONTYPES IS .836812

THE RATIO OF  BB  TO  AA  GENOTYPES IS .150865
```

Figure 24b

author ran a series of eleven runs wherein the following initial conditions were imposed on each run:

(a) the initial genotypic distribution was 1:2:1,
(b) the number of offspring was 1000,
(c) the number of generations was 5,
(d) there was no selection against any female parent genotypes; hence F1=0, F2=0, and F3=0,
(e) all offspring survived; hence C1=1, C2=1 and C3=1, and
(f) there was no selection against the male AA or the male AB parent; hence M1=0 and M2=0.

For each run, the selection against the BB male genotype varied. The first run had no selection against the male BB genotype and the last run had complete selection against the genotype. Hence, for the last run the value of M3 was 1. In the remaining runs, the selection coefficient was incremented by a value of 0.1. Thus, the 11 runs corresponded to a set of runs in which the value of M3 was 0, 0.1, 0.2, ... , 0.9, 1.0 respectively.

An analysis of these results indicated that, the run corresponding to M3=0 gave results which agreed very well with the conclusions of the Hardy-Weinberg law. This should not be surprising since for these initial conditions, the basic assumptions of the program are in accord with all of the hypotheses necessary for the validity of the law. (This statement is predicated on the idea that 1000 offspring is effectively an infinite population). As M3 was increased from 0, through the values 0.1, 0.2, etc., there was a gradual decrease in the proportion of BB genotypes measured at the fifth generation. A comparison of the proportion of BB genotypes in the population at the fifth generation for various values of M3 indicates that this proportion decreased as M3 increased. In fact, for M3=1, that is for complete selection against the mating of the BB male parent, the proportion of the entire population that consists of BB genotypes is only 0.061. This proportion is about 0.1 of the number of AA genotypes remaining in the population at the end of the fifth generation. The genotypic ratios at the end of the fifth generation for the run corresponding to complete selection against the male BB genotype were 0.571:0.368:0.061.

This example of a set of runs was presented to illustrate the many and varied experiments that may be performed with such a program. In fact, the actual performing of the computer based experiments is very easy once the program has been developed. It is the analysis and interpretation of the results that is time consuming and difficult. Because it is so easy to carry out such experiments, it is very tempting to "just go ahead and try several different sets of initial conditions and see if any interesting results are forthcoming". Needless to say, such a procedure cannot be expected to be very effective or productive. Nevertheless, the student is urged to do a bit of such experimentation just to get a feel for the capabilities and limitations of the program. One thing the student will learn from such an experience is that the analysis and interpretation of the results from several runs is not a trivial task. In fact, this task may be even more difficult than the development of the program which produced the results. This should not be surprising because the program runs are analogous to a series of experiments and it is well-known that the analysis of experimental data is not a trivial task. The paradox of the seeming ease with which computer output data can be generated compared to the not so great ease which is required to analyze and interpret such data is part of the lore that is well known in computing centers.

Other Extensions

In this section, extensions of the previous techniques to describe other genetic phenomena are presented. The procedures for carrying out these extensions or modifications will be presented in written form only. Neither a flowchart nor a program will be presented. Nevertheless, it is hoped that the written presentation will be of sufficient detail to enable the student to readily carry out the modifications. There are no essentially new ideas involved in the extensions and thus, it was not felt necessary to give a "full blown" presentation. Time permitting, the student is urged to actually carry out some, or all, of these modifications and extensions. In doing so, however, all of the justifications for the necessary modifications should be carefully stated. The program should be run and experimented with to obtain insight into the problem.

All of the problems considered so far in this chapter were restricted to a single loci at which only two alleles resided. For many physical characteristics, it is known that it is necessary to postulate the existence of more than two alleles at a locus in order to "reasonably" explain the distribution of the different states of the characteristic in the offspring population. To illustrate a method of treating more than two alleles at a single loci, we consider the specific case of three alleles at a single loci.

Denote the alleles by A, B, and C, and for the purposes of this discussion let us suppose that the physical characteristic is the height of the offspring. In this hypothetical example, the A allele is considered the allele for tallness; the B allele, the allele for average height; and the C allele is considered the allele for shortness. We further assume that the mating population is infinite in number and that the genotypic descriptions of the male and the female parents are identical. Since there are the three alleles at the locus, the possible genotypes in the population are AA, AB, AC, BA, BB, BC, CA, CB, and CC. However, we are further assuming that there is no distinction among the genotypes and therefore the AB and BA genotypes are identical, as are the BC and CB genotypes and the AC and CA genotypes. Thus, the distinct genotypes in the population are AA, AB, AC, BB, BC and CC.

The program is developed by mimicing the mating of males and females randomly chosen from a population initially specified by the respective numbers of AA, AB, AC, BB, BC and CC genotypes. These numbers will be designated in the BASIC programming language by A1, A2, A3, B2, B3, and C3 respectively. If N designates the number of individuals in the parent population, then

$$N = A1+A2+A3+B2+B3+C3,$$

and the respective genotypic ratios are

$$A1/N,\ A2/N,\ A3/N,\ B2/N,\ B3/N,\ \text{and}\ C3/N.$$

The random choice of a parent from such a distribution is accomplished by comparing the value, RND, obtained from the random number generator to the fractions:

$$A1/N,\ (A1+A2)/N, (A1+A2+A3)/N,\ (A1+A2+A3+B2)/N,\ (A1+A2+A3+B2+B3)/N$$

and 1.0.

These fractions specify a set of points on the interval (0,1) which bound the intervals whose lengths are equal to the respective genotypic ratios. Thus, the interval containing RND will specify the desired parent genotype. This procedure is used for the selection of the male and female parent alleles. Once the parent genotype has been determined, the procedure for selecting the allele to descend to the zygote is the same as that described in preceding sections. It may be the case that, if for example, a BC parent genotype were selected, the likelihood of a B allele descending is much greater than the likelihood of a C allele descending. In this event, when selecting the allele for descent, the value 0.5 would not be the value to which the RND should be compared. Rather the RND should be compared to a value specifying the degree of preference in selecting a B allele over a C allele. We remark in passing that is not easy to experimentally obtain or determine such a value.

The tallying of the genotypes of the offspring is accomplished as in the other programs. If it is desired to grow successive generations, the offspring are assumed to be the parents of the next generation and the program modified to "grow" several generations. A program describing the mating process will be somewhat longer than the first program in this chapter; however, no essentially new ideas or procedures are required to develop the program.

Another extension of interest is the description of the genotypic distribution resulting from a population described by two alleles at each of two distinct loci. Let the pair of alleles at the first locus be denoted by A and B, and the pair of alleles at the second locus be denoted by C and D. It is again assumed that the population is infinite and that there is no distinguishing between the heterozygotes AB and BA, and the heterozygotes CD and DC. The possible genotypes in the population are:

AACC, ABCC, BBCC, AACD, ABCD, BBCD, AADD, ABDD and BBDD.

Thus, we are assuming independent assortment; that is, the pair of alleles at the locus occupied by the A or B alleles segregate independently of the pair of alleles at the locus occupied by the C or D alleles. (This restriction could be removed but the development of a meaningful set of rules governing the selection of the alleles is not an easy task).

The random selection of a genotype from a parent population consisting of nine different genotypes proceeded in the same way as described in the previous example. The genotypic ratio of the parents are specified by specifying the initial numbers of each genotype and then evaluating the corresponding genotypic ratios. These ratios are then "layed out" along the interval (0,1) and the correspondence between each interval and a parent genotype is assigned. A value of RND is selected, and the interval which contains this value designates the parent genotype. Both parent genotypes are selected in the same manner. The procedure for determining the genotypes of the offspring is the same as used in the other examples. Provision must be made for tallying a larger number of distinct genotypes. This is accomplished by the inclusion of a greater number of branching statements (IF statements) in the program. The program may be extended to describe the genetic evolution over several successive generations in a manner very analogous to that used to so extend the first program.

Of the work presented in this chapter, other extensions are also possible. Some of them are listed below:

(a) Modify each of these programs to mimic the genetic evolution of small populations.

(b) Modify the programs to include selectivity and survivability.

(c) Modify the small population program shown in figure 23a,b to include the effect of migration.

(d) Modify the small population program shown in figure 23a,b to include the effect of prescribed mating. The prescription is to be supplied by the user.

(e) Rewrite one of the programs to permit the distinction between heterozygotes, that is, between the AB and BA genotypes, etc.

(f) Develop a program to mimic the mating among parents of different ages. The specification of the requisite time periods, and the proper accounting of the phenomena associated with each time period, is a critical part of the program. The development of such a program is quite a detailed process and should probably be a term and/or class project.

Comments on the Chapter Problems

The problems at the end of this chapter suggest other computer based experiments to be performed. The problems also suggest several modifications of the programs. Some of the program modifications may require considerable time and effort in order to insure that the resultant program is thoroughly debugged and performs as desired. In some of these problems, it may be easier to develop the entire program from "scratch" rather than try to modify an existing program. The student should also note that there are several equivalent, in the sense of producing the same output, ways to modify the program. In this regard, it should be pointed out that different but equivalent programs may indeed give different output. This is due to the finite length arithmetic carried out by the computer. Because the computer performs arithmetic using only a finite number of digits, there is round-off error. Round-off error can affect equivalent algorithms differently. For example, it is possible to add a column of figures in one order and to again add the same column of figures in another order and yet get different answers. Such discrepancies are due to the lack of precision caused by using a fixed, but finite, number of digits in the arithmetic. In many of the problems, it will be of interest to run each program several times and to then compare the variability of the results. In this way, some feeling can be obtained about the sample size and the stability of the program.

PROBLEMS

CHAPTER VII

Note: In discussing the results obtained with the previous or similar programs, it will usually be necessary to make several runs with identical starting conditions. The averages of the results of these runs will then provide a norm or standard against which other runs may be compared. Because the averages so obtained are not exact, and because of the possibility of great deviation in a single run, it is not expected that accurate quantitative comparisons of results can be made. Thus, in the problems below which require a discussion of results, it is only expected that qualitative remarks be made. The purpose of the runs is to provide some insight into the genetic phenomena assuming the validity of Mendel's description of the transference of hereditary characteristics.

1. Make several runs of the program listed in figure 2. Discuss the results obtained from making:
   (a) Several runs with the same value of N, and
   (b) Several runs with different values of N.

   Is your intuition about sample size and the variation of results confirmed? Discuss.
2. Make several runs using the First Population Genetics Program listed in figure 5. Discuss the results obtained from the following variations of the starting conditions:
   (a) Hold A1, A2, A3 and G constant, vary C,
   (b) Hold A1, A2, A3 and C constant, vary G, and
   (c) Hold A2, A3, C and G constant, vary A1.

   Try other variations of the starting conditions. Discuss your results.

3. Using the program in figure 5, and properly choosing the starting conditions, make some runs whose results can be compared to the Hardy-Weinberg principle. Discuss your comparisons.
4. Carry out one of the projects suggested in the section entitled Related Projects.
5. Get the Small Population, Population Genetics Program listed in figure 18 up and running on your computer. Make some preliminary runs with different sets of input data to obtain a feel for the program and the results it produces. Using your own starting conditions, make some runs and discuss the results. Try to use the program as an experimental device for "growing" monogamous mating small populations.
6. Alter the method of selecting the offspring to permit a larger average number of offspring per mating. This can easily be accomplished by properly changing the statements in lines 500 and 510. Make some runs and compare the results with those obtained from problem 5.
7. Alter the method of selecting the number of offspring so that exactly 3 offspring are born of each mating. Discuss the results of some runs with this alteration.
8. Modify the program shown in figure 2 to:
    (a) Permit the specification of the initial genotypical distribution of both male and female parents, and
    (b) Permit random choice of mating partners.
9. Same as problem 8, only using the program in figure 5.
10. Modify the program shown in figure 5 to permit small, but random, decreasing changes of the BB genotypic ratio. In this way, the phenomena of mutation can be mimiced. Make some runs and discuss your results. State specifically the hypotheses used in your modifications.

11. Modify the program in figure 5 to permit a small constant change of the BB genotypic ratio. This change is to be compensated for by making the appropriate changes in the AB and AA genotypic ratios. In this way the phenomena of migration can be mimiced. Make some runs and discuss your results. State specifically the assumptions used in your modifications.

12. Modify the second population genetics program to permit the running of the program several times and to permit the tallying of the frequency with which a given offspring genotype occurs at a given generation. Run the program for a given initial distribution of parent genotypes and then plot the frequencies of occurrence of the genotype at the specified generation.

13. Extend the modification suggested in problem 10 above to include the tallying of the frequency of a given genotype at several specified generations. Include in this modification the automatic running of the program several times using the same starting conditions for each run. The tallying procedure should tally the total genotypes for all of the runs. The program so extended is somewhat analogous to the program developed by Schaffer for infinite mating populations. Run the program. Discuss the results so obtained. The presentation and discussion of the results may require considerable thought. Because of this, and because the program alterations are considerable, this problem should probably be a term project.

14. Extend the program described in problem 9 to include the ability to:

    (a) Run the program several times, and

    (b) Tally the male and female genotypic descriptions for several specified generation.

Choose an initial genotypic distribution and run the program several times. Plot tallies given by the program. You should note a distribution similar to that given by Schaffer.

15. Get the multi-generation, selectivity and survivability program of figure 20 up and running on your computer.
    (a) Make several runs varying only one survivability coefficient.
    (b) Make several runs varying only one selectivity coefficient.
    (c) Make up your own set of starting conditions and make several runs with these conditions.

    Discuss the results obtained from each of the above.

16. Alter the program shown in figure 23 to include the effect of a slow change in the selectivity coefficient. State clearly your basis for the rule of change that you chose. Discuss the results of some runs. Note that, if the variation of the selectivity coefficient is small, it is not easy to compare the results obtained with the modified program to the results obtained from the program depicted in figure 23. This difficulty is due to the random nature of the process and hence, of the results themselves.

17. The same as problem 16 except alter the survivability coefficient.

18. The same as problem 16 except alter the survivability and/or selectivity coefficients in such a way that either one or the other, or both, are sex dependent.

REFERENCES

CHAPTER VII

Mettler, L. E. and Gregg, T. G. (1969). Population Genetics and Evolution. Prentice-Hall, Englewood Cliffs, N. J.

Simpson, G. G., Pittendrigh, C. S. and Tiffany, L. H. (1957). Life, An Introduction to Biology. Harcourt Brace. New York, N. Y.

Srb, A. M. and Owen, R. D. (1955). General Genetics. W. H. Freeman. San Francisco, CA.

Strichberger, M. W. (1968). General Genetics. The MacMillan Co., New York, N. Y.

Volpe, E. P. (1967). Understanding Evolution. W. C. Brown Co. Dubuque, Iowa.

VIII

RANDOM PROCESSES

Introduction

As mentioned in the previous chapter the analysis of genetic phenomena is usually accomplished with the aid of classical probability theory. In contrast, the method of analysis that we have used was based upon the repeated use of the random number generator to mimic or simulate the actual flow of the genes from the parents to their offspring. This suggests the possibility of attacking other probabilistic problems in the natural sciences in a similar manner. Our analysis of genetic phenomena consisted of the following steps:

1. The establishment of a mechanism, or experiment, whereby genes were passed from the parents to the offspring.
2. The construction of a computer based experiment (program) which would mimic the mechanism.
3. The repeated use of the program to permit the determination of the distribution of the alleles in a large number of matin

In the work described below, the same procedure will be followed. Thus, we will first attempt to construct or contrive an experiment (mechanism) which describes the problem and to then repeatedly mimic th experiment with the aid of a computer. The relevant quantities will be tallied and then used to calculate the estimated probabilities. The pr sent chapter, which should be considered as introductory, will serve to illustrate some computational procedures and to comment upon their effe tiveness. The techniques will be presented by considering a variety of simple examples. For each example, an analogous real world experiment w be hypothesized and then the experiment will be mimiced on the computer.

It is hoped that such a procedure will also serve to better relate the problem to reality.

This chapter contains an appendix which presents a technique for actually calculating the desired probabilities. The technique is based upon an analysis of the computer based estimation of the probability. The presentation in this chapter will consist in a thorough discussion of several problems taken from probability. This should give the student a feeling for what can be accomplished. The technique is intuitive and, for this reason, quite easy to follow.

A Very Simple Problem

The student is familiar with the fact that the probability of throwing an ace in one toss of a die is one-sixth. Most modern probability theory texts will arrive at this result by stating that the probability of throwing a die on a single toss of a die is the ratio of the number of possible outcomes which are an ace to the total number of possible outcomes. Thus, for a single throw of the die, since the number of possible outcomes which are an ace is one and the total number of possible outcomes is six, the probability is one-sixth. It is assumed that all possible outcomes are equally likely and in the subsequent work this assumption will be made unless otherwise stated. The probability, as calculated in the above manner, uses knowledge pertaining only to the single throw of the die. Nevertheless, the phrase, "the probability of throwing an ace on a single throw of a die is one-sixth" can usefully be interpreted to mean that if the die is thrown a large number, N, of times and a tally, T, is kept of the number of aces thrown, the ratio T/N is very close to one-sixth. This interpretation arises from the intuitive notion that by throwing the die enough times, the ratio T/N can be made arbitrarily close to one-sixth. This suggests the simulating of a throw of the die with the aid of a random number generator, which in the subsequent discussion will be denoted by RNG. Repeated throws require repeated use of the RNG. The computational procedure is now quite evident and simply consists of using the RNG "to throw the die", tallying whether or not an ace is thrown, and repeating the process N times. The determination of whether or not an ace has been thrown is based upon the following two facts: (1) it is known that there are six possible outcomes on the throw of a fair die and (2) the RNG produces a random number which is uniformly distributed in the unit interval; that is any number in the interval $(0,1)$ has equal probability of being selected. This suggests that by subdividing the unit interval into six subintervals of equal length and "assigning" the first interval to the throw of an ace, it is possible to mimic the actual throw of an ace. Thus, if the RNG produces a number, R, such that $R \leq 1/6$, we will say an ace has been thrown, and if $R > 1/6$, then we will say that some other number, not an ace, has been thrown. After N repetitions, the tallied number of aces divided by N is the probability estimate. The student will recognize that this procedure is entirely analogous to that used to estimate the genotypical description of the offspring.

Your author deliberately uses the word estimate rather than calcula since the result is indeed an estimation of the numerical value of the probability as calculated in the classical manner. The procedure, which is called the frequency method, relies on the intuitive idea that if the number N, of trials is large enough the difference between the estimat value for the probability and the value for the probability as obtained in the more conventional manner should be very small. The speed of the modern high-speed digital computer is such that the mimicing of 10,000 or 100,000 trials of simple experiments can be done in seconds and thus the mimicing procedure is feasible for some probability problems. However, there are a great many "seemingly simple probability problems" whose answers cannot feasibly be obtained in this manner because the computational cost would be too great. Nevertheless, there are a number of interesting and non-trivial problems whose required probabilities can be effectively estimated in this manner. This number is sufficiently large, and the problems sufficiently diverse, so that a discussion of some of these problems should be of help and interest to the student. In addition, there is a wide range of problems which can only be analyzed in this manner and many of them are of sufficient importance so that the possible large computational cost is accommodated. These procedures are called Monte Carlo methods. They are even useful in obtaining estimates to answers to problems which are not probabilistic such as determining the area of an irregular shape, calculating the heat distribution in a bar whose ends are maintained at different fixed temperatures, etc. Finally, this method of attacking probabilistic problems provides the student a further opportunity in model building. In this way the relation between the actual and the theoretical is more clearly demonstrated.

A flowchart and a computer program for estimating the probability of throwing an ace in one throw of a die is given in figure 8.1. The program input is N, the number of times it is desired to perform the experiment of throwing the die, and the output is the estimated probability. By varying the magnitude of N the accuracy of the estimated

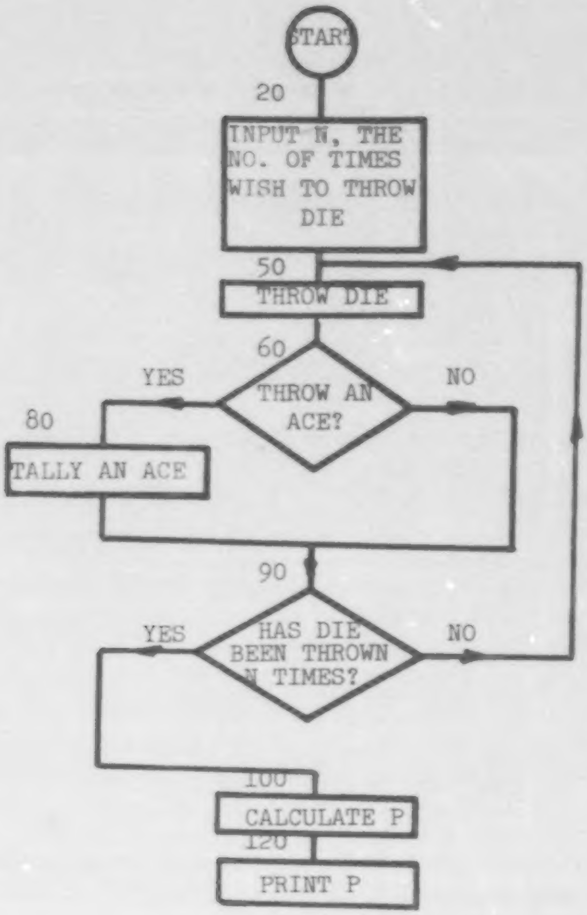

START
20
INPUT N, THE NO. OF TIMES WISH TO THROW DIE
50
THROW DIE
60
THROW AN ACE?
YES
NO
80
TALLY AN ACE
90
HAS DIE BEEN THROWN N TIMES?
YES
NO
100
CALCULATE P
120
PRINT P

Fig. 8.1

```
  1  REM ESTIMATES PROB. OF AN ACE ON SINGLE THROW OF DIE
  5  RANDOMIZE
 10  PRINT "TYPE N, THE NO. OF THROWS"
 20  INPUT N
 30  T=0
 40  FOR I=1 TO N
 50  R=RND
 60  IF R≤0.1666667 GOTO 80
 70  GOTO 90
 80  T=T+1
 90  NEXT I
100  P=T/N
110  PRINT "PROB. OF ACE ON A SINGLE DIE THROW IS"
115  PRINT
120  PRINT P
130  END
```

probability can be determined. The determination of the throw of an ace is given in line 60 of the program. Because the number 0.166667 corresponds to the fraction 1/6, the perceptive student may well ask, "haven't you used the answer to find the answer?". In fact, it appears that the exact answer has been used to find an approximate answer! The explanation of this seeming enigma is as follows. The usual technique for mimicing the throw of a die on a computer is to write

`LET R=INT(6*RND+1)` (a)

Now the operation 6*RND+1 produces a random number between 1 and 7 and taking the integer part of this number yields an integer between 1 and 6. We associate a 1 with an ace on the throw of the die and thereby mimic the throw of an ace. By using the operation of 'integer part' we are in effect subdividing the interval (1,7) into 6 equal parts, each of unit length, and by associating the interval (1,2) with the throw of an ace we are determining when an ace is thrown. Thus, we are associating 1/6 of the interval (1,7) with an ace. Therefore, the process defined by equation (a) above actually expands the unit interval, partitions it into 6 equal parts and then assigns one of these parts a desired interpretation. Now the same end can be accomplished by directly subdividing the unit interval into 6 subinterval parts and then associating the first of these subintervals, (0,1/6), with the throw of an ace. This procedure requires less computational effort and is much easier to implement. Hence we are using it.

It is certainly true that we have ended up building the answer into our problem, albeit unwittingly. However, this turns out to be a blessing rather than a curse since it permits an easy recognition of our procedure with the approach used in classical probability theory.

In fact, if we associate the length of the unit interval with a large number, N, of throws we see that the interval (0,1/6) corresponds to N/6 throws and hence the ratio (N/6)/N=1/6 is the desired probability. This interpretation is very fruitful and we shall pursue it more in the first appendix to this chapter.

In order that the student may better appreciate the above discussion we consider the random selection of the parental genotypes in the genetics problem in a similar manner. In the simple version of that problem the number of AA, AB and BB parent genotypes was given and it was required to randomly select a parent from such a prescribed genotypical distribution. We will illustrate the procedure for selecting a parent genotype by considering a specific example. Assume that there are 200 AA, 500 AB and 300 BB parent genotypes respectively. Since there is a total of 1,000 parents, it is necessar to construct a process for selecting a random number between 1 and 1,000. The instruction

```
LET R=INT(1000*RND+1)
```

accomplishes this. This instruction expands the unit interval to an interval 1000 units in length and then subdivides this interval into 1000 equal intervals of unit length. By associating values of R from 1 to 200 with AA parent genotypes, values of R from 201 to 700 with AB parent genotypes and values of R from 701 to 1000 with BB genotypes we have constructed a process which randomly selects an AA, AB or BB parent genotype. The procedure of comparing a random number in the unit interval with the respective fractions 0.2, 0.7 and 1.0 accomplishes the same objective and is easier to implement and more economical to use. In effect we are saying that 0.2 of the parents are AA genotypes, 0.5 of the parents are AB genotypes and 0.3 of the parents are BB genotypes. This seems a more natural description.

In our die problem, as the number of throws increases, it will be noted that the estimated probability does indeed approach 1/6. However, the speed of this approach is very slow, that is about 100 times as many trials are needed to get a single decimal digit more of accuracy in the estimated probability. This result is in accordance with the well known fact (at least around computer centers) that Monte Carlo or RNG based procedures require a very large number of trials to obtain reasonable accuracy. It can be shown that the accuracy varies as the square root of the number of trials and thus an increase in accuracy of two decimal digits, that is an increase in accuracy of 100, will require $(100)^2$ or 10,000 times as many trials. Despite this very real difficulty in the obtaining of accurate answers, the method is quite useful for obtaining reasonable or crude estimates. The student should run the simple dice program for different values of N. He will note that for small values of N, like 12 or so, the results vary dramatically. This is in accord with the well known observation that much more frequently than one would intuitively expect, it is the case that the number of aces recorded in 12 throws of a die is not 2. A repetition of the experiment of throwing the die 100 times indicates less relative variation in the estimated probabilities and repeating the experiment with 10,000 throws of the die reveals approximately one tenth the previous relative variation. Such comparisons also illustrate the difficulty in obtaining accurate assessments from small samples.

It is natural to ask, "In view of this uncertainty about the answer, how does one know when a prescribed degree of accuracy has been attained?" The answer is "One cannot actually know for certain if a prescribed degree of accuracy has been attained". Theory tells us that as the number of experiments increases it becomes more and more certain that the ratio of the tallied successful experiments to the total number of experiments approaches the true probability. Your author deliberately chose the word approaches rather than the word converges because the latter word has a very definite meaning in mathematics. In mathematics, if a sequence of numbers is said to converge to a number, it is understood that by "going out far enough in the sequence" it is definitely possible to find a term in the sequence of numbers, such that this term and all others succeeding or beyond it in the sequence, are arbitrarily close to the number. In contrast, the sequence of probability estimates generated by repeating the experiment for successively larger numbers of trials N, cannot be said to converge to the probability in the sense that it is possible to find some finite value for N such that for all numbers of experiments greater than N the value of the probability as calculated would be arbitrarily close to the true probability. The key word here is the word 'all'. What is true, and can be shown, is that for a large enough value of N, the probability as estimated for larger and larger numbers of experiments, would have less and less of a chance of differing from the true probability by an arbitrarily small amount. In other words, the certainty of getting arbitrarily close to the true probability increases as the number of experiments increases. A rigorous discussion of this topic is presented in texts on mathematical probability. see for example, Uspensky.

In light of the previous discussion, the accuracy of the results is subject to question. Nevertheless, the "probability" of the methods producing very incorrect estimates is very very small providing that the number of experiments is large enough. We now present a heuristic discussion of how to probably obtain a prescribed degree of accuracy using the mimicing technique. Denote the number of repetitions of N trials by M where M may be 3 or 4. A crude method for estimating the accuracy of the result consists of M repetitions of N trials and noting the degree of agreement of the left most figures of the probability estimates obtained from each repetition. If the first 3 figures, say, are the same, then we use these first 3 figures as our probability estimate. An obvious improvement on this method is to make more effective use of the data obtained in the M repetitions of the N trials. By adding the total number of favorable experiments in all M repetitions and then dividing this sum by the total number of trials, M*N, it is possible to obtain a probability estimate corresponding to (M-1)*N more mimiced experiments with very little extra effort. Another measure of the accuracy can be obtained by averaging the probability estimates calculated from each of the M repetitions and then comparing this result to the others. If the probability estimates obtained from each of the M repetitions are considerably different, one from another, it may be necessary to increase the number of trials by a factor of 10 or 100 (hopefully not more unless computational time is of no concern) and to then rerun the programs. "Proper" comparison of the results so obtained with the earlier results should then enable a reasonable estimate of the probability to at least a few significant figures. (Note all of the hedge words.)

A Second Simple Problem

A farmer has a herd of 20 ranch cows which simultaneously become infected with hoof and mouth disease and in four weeks 15 of them have recovered while the remaining 5 have not yet recovered. Five animals are selected at random from the entire herd and it is desired to know the probability that:

(a) None of the 5 have recovered in four weeks.

(b) Exactly 3 of the 5 have recovered in four weeks.

(c) All 5 have recovered in four weeks.

The first task in the procedure for estimating the probabilities is to construct a "real world" experiment which, if repeated often enough and the appropriate results tallied, would enable the calculation of the desired probability estimates. In this example a realization of what is required in order to estimate the probabilities will suggest the experiment to be performed. The frequency method suggests that we imagine that there exists a large number N, of such herds and that from each of these herds 5 animals are selected at random. After each selection of 5 animals, the number of recovered animals is tallied and running totals of the results are kept. An estimate of the probability that all 5 have recovered is given by the ratio of the number of selections resulting in all 5 animals having recovere divided by N. Similar ratios give the required probability estimates for parts (b) and (c). An equally appropriate hypothetical experiment consists in randomly selecting five cows, tallying the health status of each cow, then replacing the five cows. The cows are allowed to mill and mix together and five cows are again selected at random and the health status is again tallied. The process is repeated N times.

The development of the computer program requires a method for mimicing the selection of 5 cows at random from a herd of 20 cows. This will be accomplished by a slight modification of the technique used to mimic the random selection of a parental genotype from a prescribed distribution of parental genotypes. The mimicing of the random selection of the first cow is based upon the fact that the proportion of recovered cows is 15/20. Recalling that the RNG produces a random number, R, between 0 and 1, we generate an R and compare it to the proportion 15/20. If $R \leq 15/20$ we say that we have selected a cow which recovered; if $R > 15/20$ we say the selected cow has not recovered. The selection of the second cow is based upon the fact that after the selection of the first cow, the herd consists of only 19 animals, and if it is supposed that the first cow selected has recovered, the remaining proportion of recovered animals would then be 14/19. The mimicing of the selection of a second cow is accomplished by comparing a second random number with the ratio 14/19 to determine the state of health of the animal. In a similar manner, if a second recovered cow has been chosen, the remaining proportion of recovered cows is 13/18 and a third R is chosen and compared to the ratio 13/18 to determine whether or not the third selected cow has recovered. On the other hand, if the initial R had been such that $R > 15/20$, we would have said that a sick cow had been selected. Since 15/19 of the remaining herd has recovered, the next value of R is compared to 15/19 to determine whether or not a recovered cow has been selected. If a recovered cow is then selected, a third value of R is compared to the fraction 14/18 to determine the status of the health of the third cow drawn. The procedure should now be clear and the process is repeated two more times, thus mimicing the choosing of 5 cows at random. Of course, for the calculation of part (a) if ever an unhealthy cow is selected there is no need to continue mimicing the experiment since only those experiments resulting in the selection of 5 recovered cows are to be tallied along with the total number of experiments. The implication of this fact should be included in the program to minimize computer time. In the program shown below this was not done.

The selection procedure has been discussed in such great detail to emphasize the closeness of the computer based experiment to the actual experiment. We again repeat, the procedure is to mimic, with the computer, the way in which the probabilities would actually be estimated if they were to be estimated experimentally. In a real sense we are constructing computer based experiments. It is usually the case that the most difficult part of the problem is the imagining of the hypothetical 'real world' experiment which, if actually carried out, would yield the data necessary to estimate the desired probabilities. In this regard, as an aid in the imagining of the experiment, it may be of assistance to disregard the cost or practicality of the proposed experiment; just assume that whatever the procedure is; it could be carried out regardless of cost, size, manpower, etc. The sole criteria to be met by the hypothetical experiment is that if it were carried out it would yield data, which if tallied, could be used to form the ratios which are to be the estimates of the desired probabilities. The construction of the computer program which mimics the experiment is usually quite evident once the 'real world' experiment has been carefully delineated.

A computer program which mimics the experiment is shown in figure 8.2 and the procedure for selecting the 5 cows at random is accomplished in instructions 120-170. In particular, instruction 140 is the mechanism whereby the proportions are altered each time a recovered cow is chosen. The initial proportion of recovered cows is specified

```
10 REM     A SECOND SIMPLE PROBLEM
20 RANDOMIZE
30 PRINT "TYPE  N,  THE NUMBER OF EXPERIMENTS     ";
40 INPUT N
70 LET S=0
80 LET U=15
90 LET V=21
100 FOR I=1 TO N
110 LET J=0
120 FOR K=1 TO 5
130 LET R=RND(0)
140 IF R<=(U-J)/(V-K)GO TO 160
150 GO TO 170
160 LET J=J+1
170 NEXT K
180 IF J=5GO TO 200
190 GO TO 210
200 LET S=S+1
210 NEXT I
220 LET P=S/N
230 PRINT "THE  PROBABILITY IS";P
240 END

READY
```

Fig. 8.2

in instructions 80 and 90. By altering these data, other proportion may be examined. Line 160 in the program tallies the number of recovered cows out of the 5 that are selected and line 200 tallies the number of times 5 recovered cows are selected. The probability i calculated in line 220. By changing instruction 180 to read

```
180   IF J=3 GOTO 200
```

part (b) can be answered and by changing the number 3 in the above line to the number 0, part (c) can be answered.

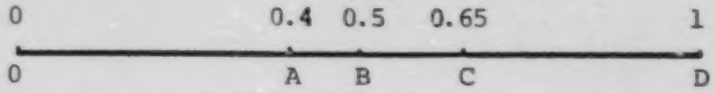
A Third Problem

A geneticist has a large population of experimental white mice. 50 percent of the mice have short tails, 25 percent have discolored eyes and 40 percent of the mice with discolored eyes also have short tails.

(a) What is the probability that a mouse selected at random has discolored eyes and also has a short tail?

(b) What is the probability that a mouse chosen at random has neither of the afflictions?

The "real world" experiment would consist in choosing a mouse at random and then recording whether it had a short tail, discolored eyes or both. The experiment would be repeated a large number of times and a tally made of the number of mice that had the respective afflictions. In order to computationally mimic the random drawing of a large number of mice and to then determine the respective affliction of each of the mice so selected, it is first necessary that the proportions of mice having the respective afflictions be designated proper on the interval (0,1). The designation of the proportion of mice havi short tails is accomplished by letting the interval (0,0.5) correspon to short tailed mice. Since 40 percent of the mice with discolored eyes also have short tails, we see that 40 percent of 25 percent, or 10 percent, of the total mice population have both afflictions. Thus, the interval (0.4,0.5) corresponds to the proportion of the mice population having both afflictions and the interval (0.5,0.65) corresponds to the proportion of mice having only a discoloration of the eye. A pictorial representation of the distribution of the proportions of the afflictions in the unit interval is



where:

i) segment 0B designates the short tail mice,

ii) segment AC designates the discolored eye mice,

iii) segment AB designates the mice with both afflictions, and

iv) segment CD designates the unafflicted mice.

These segments are the proportions of the mice population having the respective afflictions. The selection of a mouse at random and the determination of its affliction is accomplished by choosing a random number and then noting in which interval the random number falls. Because the population is "large" these proportions will not change regardless of the number of mice selected from the population. The construction of the program for part (a) is now straightforward and is represented in the flowchart, figure 8.3.

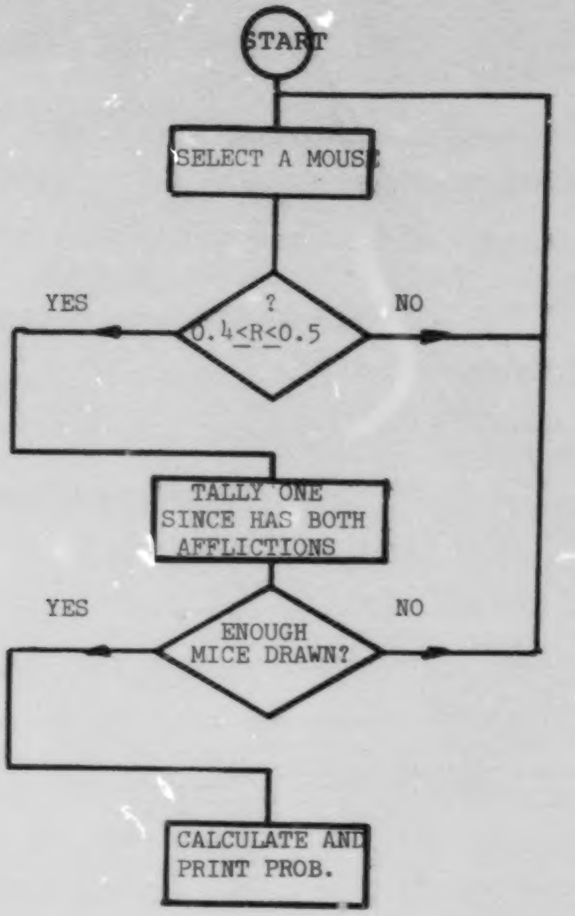
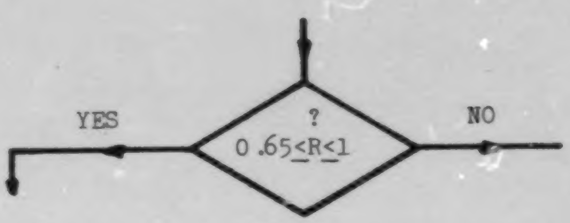

START
SELECT A MOUSE
YES
?
0.4≤R≤0.5
NO
TALLY ONE
SINCE HAS BOTH
AFFLICTIONS
YES
ENOUGH
MICE DRAWN?
NO
CALCULATE AND
PRINT PROB.

Fig. 8.3

Flowchart for Third Problem

387

The flowchart of the program to mimic the experiment to obtain the tallies necessary to answer part (b) is obtained by replacing the first test by



Those students who have had the rudiments of probability theory will recall that the answers are readily obtainable from the pictorial representation, on the unit interval, of the respective afflicted proportions of the mouse population. Thus, it may seem that by attacking the problem in this manner, we have used a megaton to only partially shoot down a fly. However, the construction of such computer based experiments is an art. The capability to readily construct such programs is very helpful when these programs are to be integral parts of larger programs. Thus, your author feels that the more examples you have to look at, the easier it is for you to acquire the art.

A Fourth Problem

A biologist has two cages containing 500 and 1000 flies respectively. Because of unsanitary conditions in the laboratory 1/5 of the flies in the first cage become sick and 1/4 of the flies in the second cage become sick.

(a) What is the probability that a fly chosen at random comes from the second box and is ill?

(b) If 30 percent of the flies in the second cage have a wing mutation what is the probability that a randomly selected fly is not ill but does have the wing mutation?

An experiment whose repetition would produce data permitting the estimation of the required probabilities could consist in randomly selecting a fly from either cage, noting its health status and the cage it came from and then returning the fly to its original cage. The estimation of the answer to part (a) will be obtained by mimicing the random selection of a fly, tallying whether it came from the second cage and was ill, and then repeating this procedure a large number of times. The determination of the cage from which the fly was selected is made by comparing a random number R1, to the ratio 500/1500 or 1/3. In discussing this problem, and others to follow, R1, R2, R3, ,,, etc. will denote random numbers in the interval (0,1) as successively produced by the RNG. If $R1 \geq 1/3$, the fly is said to come from the second cage. Given that the fly came from the second cage, a second random number R2, is chosen and compared to the ratio 1/4 and if $R2 \leq 1/4$, the fly is said to be ill. The development of the program should now be evident.

The answer to part (b) requires the determination of whether or not the randomly selected fly came from the second cage, had a wing mutation and was not ill. To accomplish this we select R3 and, as above, if $R3 \geq 1/3$ the fly is said to come from the second cage. By comparing R4 to 3/10, the determination of whether or not the fly has a wing mutation can be made and by comparing R5 to the ratio 1/4 the health of the fly can be ascertained. Thus, if the three inequaliti $R3 \geq 1/3$, $R4 \leq 3/10$ and $R5 \geq 1/4$ are all simultaneously satisfied, the fl will be said to have come from the second cage, have a wing mutation and be in good health.

```
1 REM   PROBLEM NO. 4 OF CHAPTER VIII
20 PRINT "TYPE N, THE NO. OF DESIRED EXPERIMENTS"
30 INPUT N
90 RANDOMIZE
100 LET S=0
110 FOR I=1 TO N
120 LET R=RND
130 IF R>=.333333GO TO 150
140 GO TO 210
150 LET R=RND
160 IF R<=.3GO TO 180
170 GO TO 210
180 LET R=RND
190 IF R<=.25GO TO 210
200 LET S=S+1
210 NEXT I
220 LET P=S/N
225 PRINT
230 PRINT "THE ESTIMATED PROBABILITY IS"
240 PRINT P
250 END

READY
```

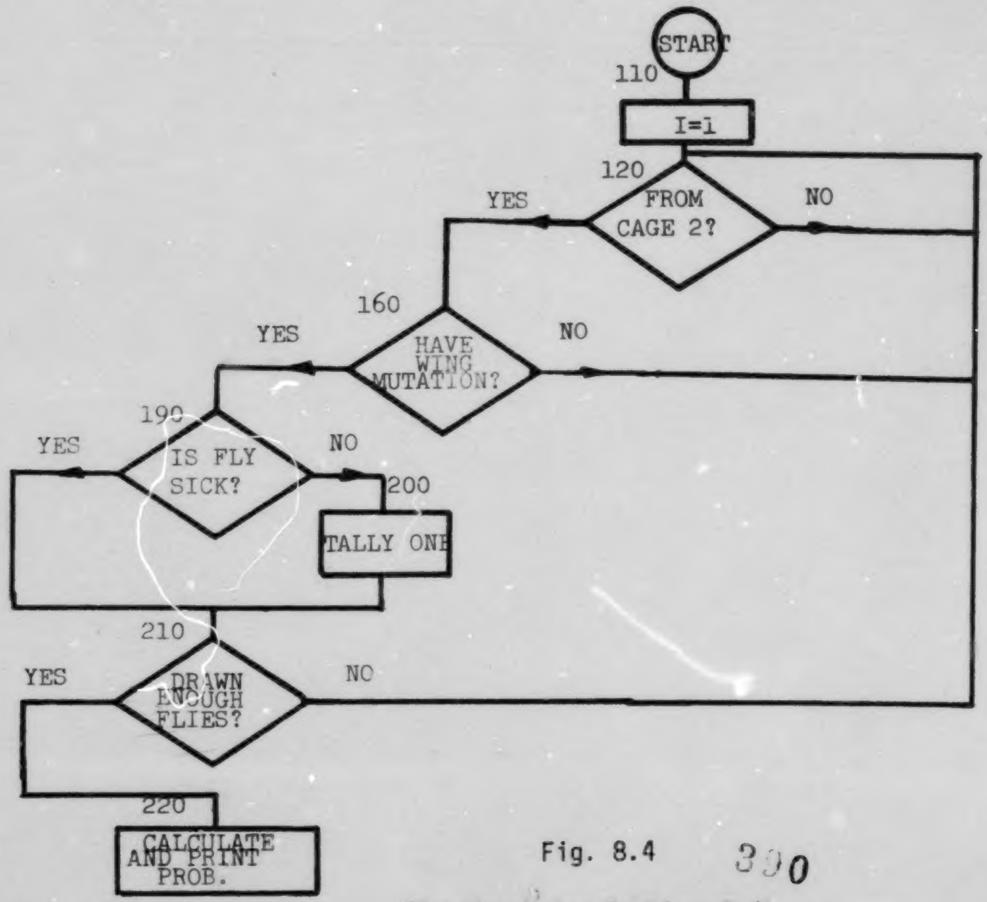

START
110
I=1
120
FROM CAGE 2?
YES
NO
160
HAVE WING MUTATION?
YES
NO
190
IS FLY SICK?
YES
NO
200
TALLY ONE
210
DRAWN ENOUGH FLIES?
YES
NO
220
CALCULATE AND PRINT PROB.

Fig. 8.4

Flowchart for Problem 8.4b

The computer program for part (b) is listed in figure 8.4 and the determination of which cage the fly is selected from is accomplished in lines 120 and 130. Lines 150 and 160 determine whether or not the fly has a wing mutation and lines 180 and 190 determine the health of the fly. The tallying is done in line 200.

A Fifth Problem

Three large cities C1, C2, and C3 experience a simultaneous epidemic of both influenza and measles. In city C1, the proportion of people having influenza is 1/8 and the proportion having measles is 1/16 while in city C2 the respective proportions are 1/10 and 3/10, and finally in city C3 the respective proportions are 1/25 and 1/5. It is assumed that all individuals are equally likely to have either disease or both. The populations of the cities are such that city C1 is twice as large as C2 and three times as large as C3.

(a) What is the probability that a person selected at random fro one of the three cities has only measles?

(b) What is the probability that a person selected at random fro one of the three cities has both influenza and measles?

(c) A person is selected at random from each city. What is the probability that exactly one of these people has only one of the illnesses?

(d) If a person is selected at random from one of the three cities, and has both diseases, what is the probability that the selected individual came from city C2?

For parts (a), (b) and (d) the experiment will consist in randomly selecting an individual from one of the three cities, and then based upon the epidemic data of that city, determining the state of the health of the chosen individual. In order to use the RNG to select an individual from one of the three cities it is first necessary to properly subdivide the unit interval in accordance with the relative proportions of the populations of the three cities. This may be accomplished by letting the entire unit interval represent the total population of the three cities and then determining that part of the interval each city should occupy. A little thought suggests that the required subdivision is obtained by designating the first 6/11 of the unit interval as representing city C1, the second 3/11 of the interval as representing city C2 and the last 2/11 as representing city C3. Thus, the choice of which city the randomly chosen individual comes from is given by the three possibilities:

(a) if $R1 \leq 6/11$ the individual is from city C1,
(b) if $6/11 < R1 \leq 9/11$ the individual comes from city C2, and
(c) if $9/11 < R1 \leq 1.0$ the individual comes from city C3.

The development of the computer program to answer part (a) is explained below and in the flowcharts shown in figure 8.5a, 8.5b and 8.5c on pages 8.26, 8.27, and 8.28.

Instruction 120 of figure 8.5a determines if the randomly selected individual comes from city C1 and if he does, line 140 ascertains whether or not he has influenza. Assuming that he does hav influenza, a marker, C, is set equal to one (line 170) to denote this, and the number of individuals with influenza is increased by one (line 180). If the individual does not have influenza the marker is set equal to zero. The marker provides a method for noting whether or not an individual has influenza and this capability is used in instruction 230 to separate those individuals having only measles from those that have both measles and influenza. Instruction 210 decides whether or not the individual has measles. Counters have been inserted at the proper places to provide for the counting of the number of individuals having the various diseases. Counters F1 and M1 count the number of individuals from city C1 who have only influenza or only measles respectively and counter B1 counts the number of individuals having both diseases. Similarly in figures 8.5b and 8.5c counters F2, M2, B2 and F3, M3, B3 count the respective number of such individuals in cities C2 and C3. The estimated probability for part (b) is obtained by replacing line 810 by

```
810  LET P=(B1+B2+B3)/N
```

and rewriting line 820 as

```
820 PRINT "THE PROBABILITY OF HAVING BOTH MEASLES AND INFLUENZA IS"
```

Part (c) requires the mimicing of the experiment which consists of the random selection of 3 individuals, one from each city, and the determination of the health status of each. This experiment is repeated a large number of times and all occurrences in which exactly one of the three selected individuals has only one disease are tallied. The ratio of the number so tallied with the number of experiments provides an estimate of the required probability.

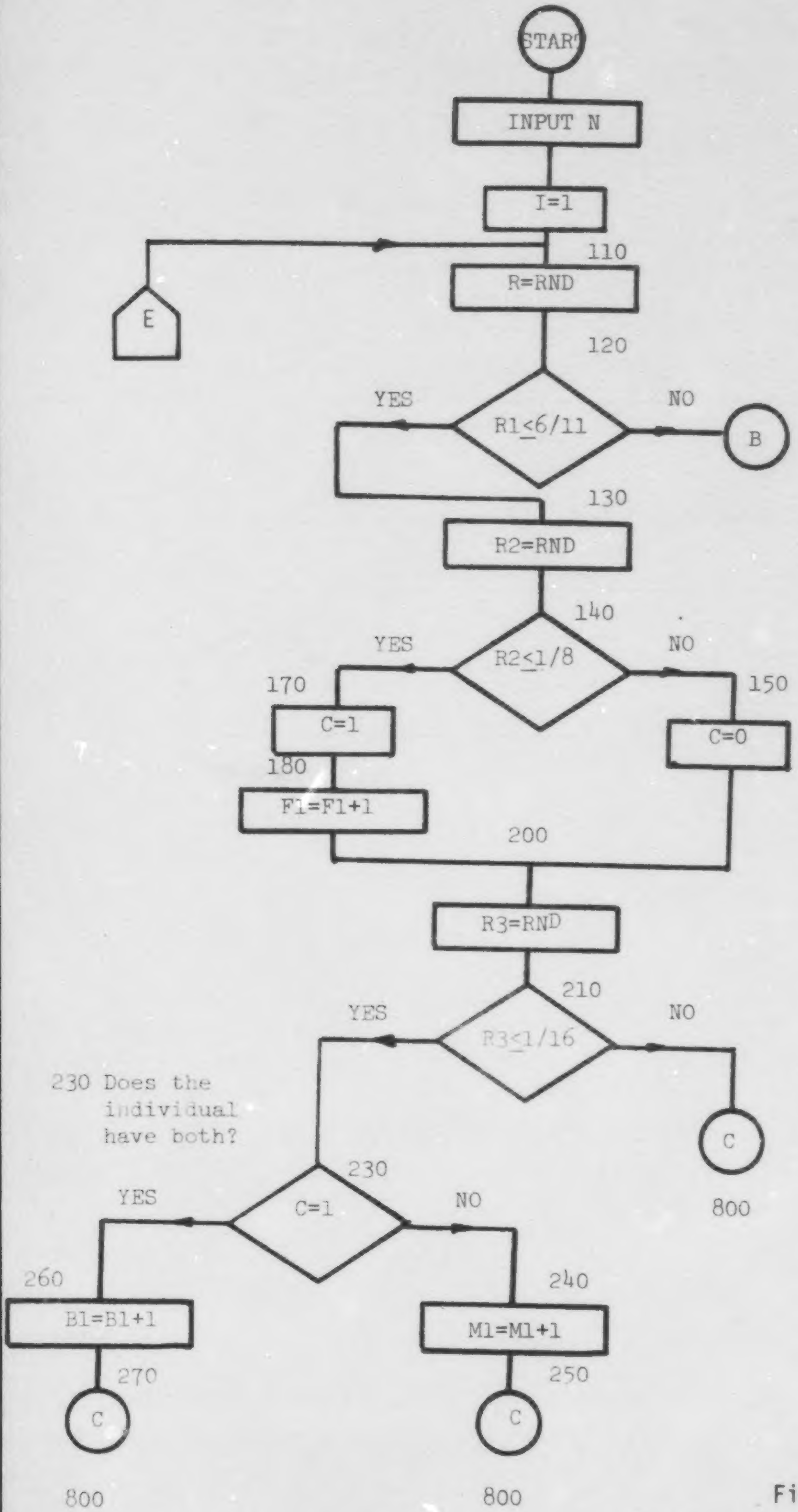
The previous program needs only slight modifications to obtain a program which will permit an estimation of the answer to part (c) Because an individual is selected from each city, the determination of which city the individual originates from can be omitted and therefore, lines 110, 120, 320 and 330 may be deleted. The determination of the health status of a randomly selected individual from a given city is accomplished by comparing R, a random number produced by the RNG, to the appropriate proportion of sick individuals in the given city. This selection must be repeated three times; each time with a different random number. For each individual the health status is recorded and a determination is made of whether or not exactly one of the three individuals has only one of the diseases. If this is so, a tally is recorded. The experiment is then repeated N times regardless of whether or not exactly one of the individuals has only one of the diseases. The quotient of the tally and N yields the desired probability estimate. To obtain an answer to part (d) it is first necessary to understand what is called for. In terms of our frequency interpretation of probability we see that this question is equivalent to asking what proportion of all the individuals having both diseases is the number of C2 individuals having both diseases.

The modification of the computer program to obtain an estimate to the probability asked for in part (d) is quite simple. This is because provision has already been made for tallying the number of individuals in each city having both diseases (lines 260, 460 and 660). Since the probability is estimated by the ratio of the number o city C2 individuals having both diseases to the total number of individuals having both diseases it is only necessary to alter line 810 to read:

```
810  LET P=B2/(B1+B2+B3)
```

and to rewrite line 820 as:

```
820  PRINT "PROB. THAT AN INDIVIDUAL HAVING BOTH DISEASES IS FROM C2
```


START
INPUT N
I=1
110
R=RND
E
120
YES
R1≤6/11
NO
B
130
R2=RND
140
YES
R2≤1/8
NO
170
C=1
150
C=0
180
F1=F1+1
200
R3=RND
210
YES
R3≤1/16
NO
230 Does the individual have both?
C
800
230
YES
C=1
NO
260
B1=B1+1
240
M1=M1+1
270
C
250
C
800
800

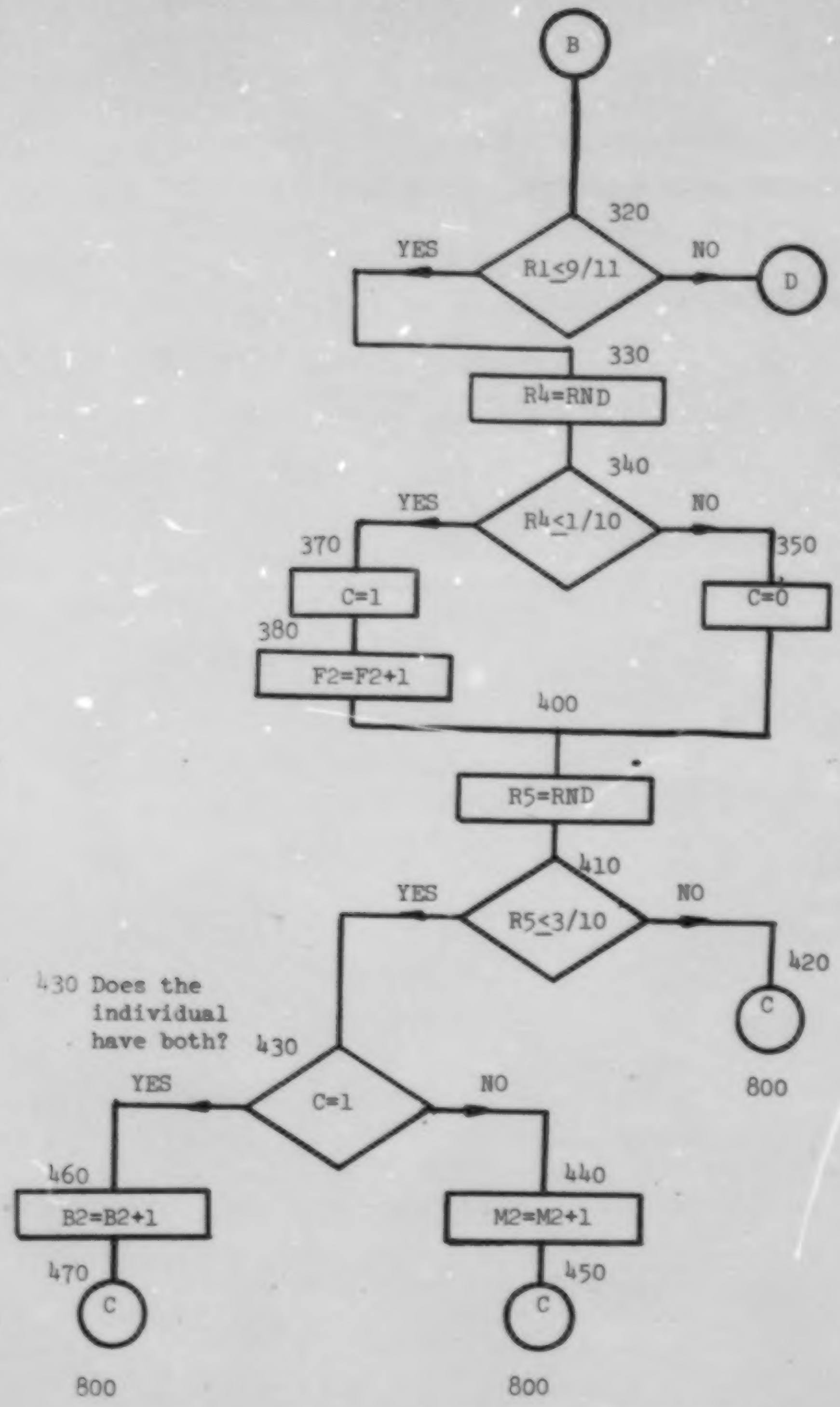
120 Is the individual from city C1?

140 Does the individual have influenza?

210 Does the individual have measles?

Flowchart for Problem #5

Fig. 8.5a


B
320
YES
R1≤9/11
NO
D
330
R4=RND
340
YES
R4≤1/10
NO
370
350
C=1
C=0
380
F2=F2+1
400
R5=RND
410
YES
R5≤3/10
NO
420
C
800
430
YES
C=1
NO
460
440
B2=B2+1
M2=M2+1
470
450
C
C
800
800

320 Is the individual from city C2?

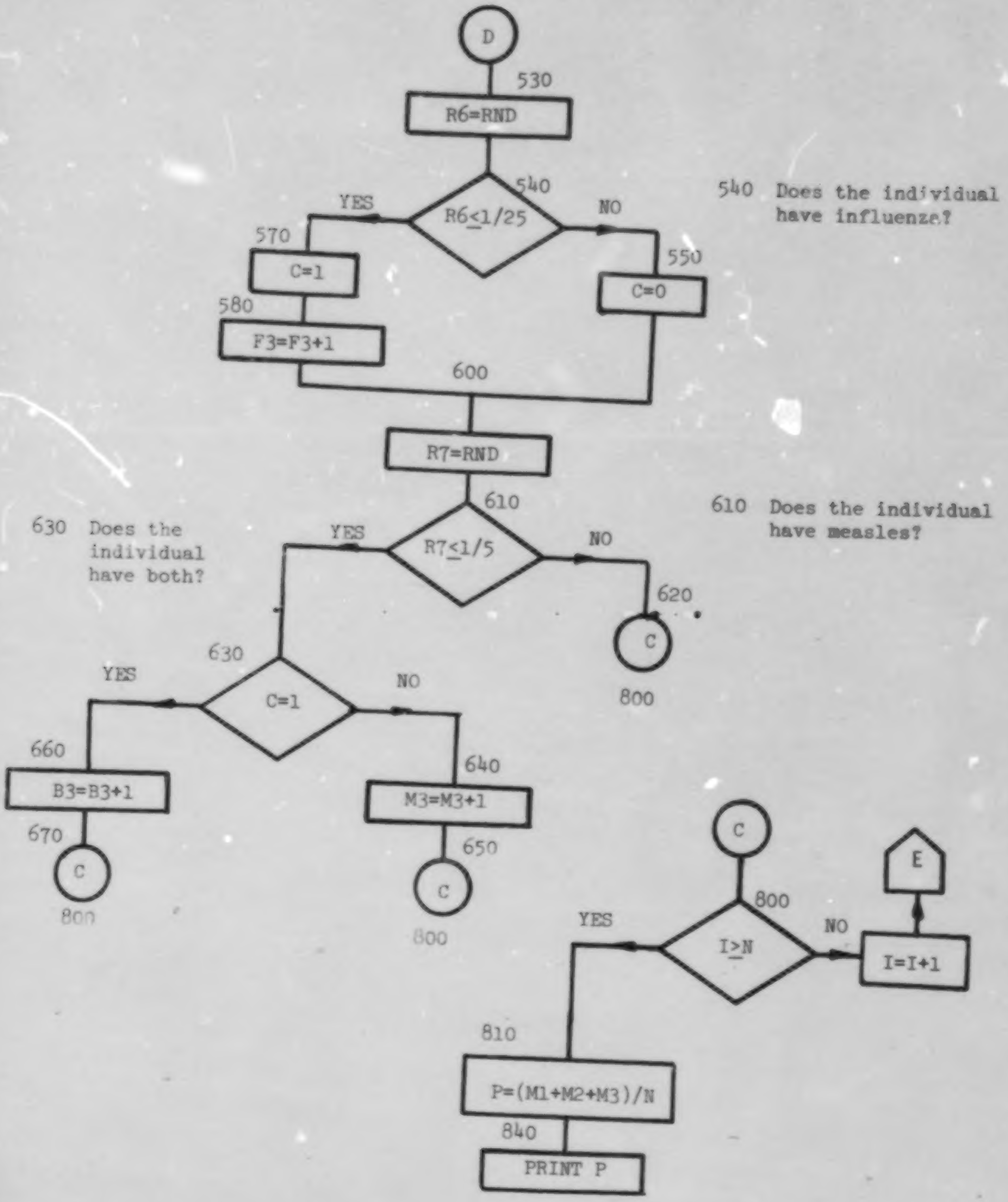
340 Does the individual have influenza?

410 Does the individual have measles?

430 Does the individual have both?

Flowchart for Problem #5 (continued)

Fig. 8.5b


D
530
R6=RND
540
R6≤1/25
YES
NO
540 Does the individual have influenza?
570
C=1
580
F3=F3+1
550
C=0
600
R7=RND
610
R7≤1/5
610 Does the individual have measles?
630 Does the individual have both?
YES
NO
620
C
800
630
C=1
YES
NO
660
B3=B3+1
670
C
800
640
M3=M3+1
650
C
800
C
800
I≥N
YES
NO
I=I+1
E
810
P=(M1+M2+M3)/N
840
PRINT P

Flowchart for Problem #5 (continued)

Fig. 8.5c

```
RBP20

1 REM   THIS IS THE FIFTH PROBLEM
2 REM
3 REM
30 PRINT "TYPE N, THE NO. OF TIMES WISH TO DO EXPERIMENT"
35 INPUT N
40 PRINT
44 REM
45 REM       LINES 50-72 INITIALZE THE COUNTERS
46 REM
50 LET F1=0
51 LET F2=0
52 LET F3=0
60 LET B1=0
61 LET B2=0
62 LET B3=0
70 LET M1=0
71 LET M2=0
72 LET M3=0
90 RANDOMIZE
100 FOR I=1 TO N
110 LET R1=RND
120 IF R1<=.545454GO TO 130
122 GO TO 320
124 REM
125 REM INSTRS. 130 TO 260 ESTABLISH HEALTH OF IND. FROM FIRST CITY
126 REM
130 LET R2=RND
140 IF R2<=.125GO TO 170
150 LET C=0
160 GO TO 200
170 LET C=1
180 LET F1=F1+1
200 LET R3=RND
210 IF R3<=.0625GO TO 230
220 GO TO 800
230 IF C=1GO TO 260
240 LET M1=M1+1
250 GO TO 800
260 LET B1=B1+1
270 GO TO 800
320 IF R1<=.818181GO TO 330
322 GO TO 530
324 REM
325 REM INSTRS. 330 TO 460 ESTABLISH HEALTH OF IND. FROM SECOND CITY
326 REM
330 LET R4=RND
340 IF R4<=.1GO TO 370
350 LET C=0
360 GO TO 400
370 LET C=1
```

Fifth Problem

```
380 LET F2=F2+1
400 LET R5=RND
410 IF R5<=.3GO TO 430
420 GO TO 800
430 IF C=1GO TO 460
440 LET M2=M2+1
450 GO TO 800
460 LET B2=B2+1
470 GO TO 800
519 REM
520 REM  INSTRS. 530 TO 660 ESTABLISH HEALTH OF IND. FROM THIRD CITY
521 REM
530 LET R6=RND
540 IF R6<=.04GO TO 570
550 LET C=0
560 GO TO 600
570 LET C=1
580 LET F3=F3+1
600 LET R7=RND
610 IF R7<=.2GO TO 630
620 GO TO 800
630 IF C=1GO TO 660
640 LET M3=M3+1
650 GO TO 800
660 LET B3=B3+1
670 GO TO 800
800 NEXT I
810 LET P=(M1+M2+M3)/N
820 PRINT "THE PROB. OF AN IND. HAVING ONLY MEASLES IS"
830 PRINT
840 PRINT P
850 END
```

Fifth Problem (continued)

A Sixth Problem

In three large southern pine forests, some of the trees are infected with a stem disease. One-third of the trees in the first forest are infected, one-fourth of the trees in the second and one-fifth of the trees in the third forest are infected. A tree is selected at random from each forest:

(a) What is the probability that exactly one tree has the disease?

(b) If exactly one of the trees selected has the infection, what is the probability that the infected tree came from the second forest?

The experiment for part (a) is readily imagined. It consists in merely drawing a tree from each forest and tallying one marker if only one of the selected trees is infected. The trees are then "replaced" and the experiment is repeated. The computer based mimicing of the experiment consists in comparing three different random numbers to 1/3, 1/4 and 1/5 respectively; and, whenever only one of the R's is less than its respective ratio, a number is tallied.

The obtaining of an answer to part (b) is not so straightforward since it is not readily evident what data should be tallied. If probability is thought of as a proportion or a ratio, it is recognized that the question in part (b) is equivalent to asking what proportion of the total number of selections resulting in exactly one infected tree is the number of selections resulting in exactly one infected tree coming from the second forest? The experiment is thus the same experiment used to answer part (a), only an additional record is kept designating whi forest produced the infected tree. As the experiment is repeated, tallies are kept of the number of times the infected tree comes from each forest whenever only one infected tree is recorded among all three of the trees selected. The proportion of the number of tallies of exactly one infected tree coming from the second forest to the total number of experiments in which exactly one infected tree is selected from all three forests is the desired probability. The computer program to answer part (a) would only need to be slightly altered to count the number of times that the selected tree comes from a given forest and is also the only infected tree selected from the three.

We will not develop the computer program; that task will be left to the student as a problem. However, it is necessary to point out that the procedure described above for estimating the probability could be very expensive of computing time. This is because the procedure depends upon generating large sets of data from data which is itself a subset of generated data. In this example it is necessary to generate a set of data consisting of a large number of experiments in which only one tree was infected of the three randomly selected trees. This set must be large enough so that it contains a reasonable representation of infected trees drawn from each forest. Since the procedure for determining an experiment in which exactly one tree is infected may produce many experiments in which no trees are infected or in which 2 or 3 trees are infected, it is seen that the building up of a large number of desired experiments could require considerable computing time. For example, suppose that only one-thousandth of the trees in each forest were infected. Then, given that exactly one of the chosen trees was infected, it is evident that the probability that the infected tree came from the second forest is 1/3. However, the generation of a large number of trials in which exactly one tree of the three randomly selected trees was infected would require many many selections of groups of three trees. The excessive number of selections would be required because nearly all of the triple of trials would result in trials in which no tree was infected. This type of computational difficulty is frequently encountered when using a computer to estimate probabilities by mimicing an experiment. Frequently, alternative but valid, experiments can be constructed which can be more economically mimiced on the computer. However, the construction of such experiments may require considerable ingenuity.

As an example of a problem for which experiments can be imagined and which require different amounts of computational effort we consider the problem of determining the probability of a three-nothing split in trump in a bridge hand, given that the declarers have ten trump between them. For those who are unfamiliar with bridge we state the problem in more familiar terms. All of the cards of a 52-card deck are randomly dealt out to form four distinct hands of 13 cards each. It is given that two of the hands contain a total of 10 cards in one suit. What is the probability that the remaining three cards of the suit all lie in only one of the two remaining hands?

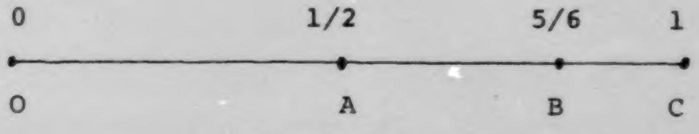
At first sight it would seem that the experiment would be to deal out four hands and then to note if there was a total of 10 cards in one suit in two of the hands. If this was found to be the case, then this fact would be recorded and it would further be recorded whether or not the three remaining cards of that suit were all in one of the other two hands. The drawback to such a procedure is obvious; it might be necessary to deal several sets of four hands before a set of four hands was dealt which had the required 10 cards of the single suit in two hands. Furthermore, the mimicing of the dealing of 52 cards requires at least 51 uses of the random number generator together with a large number of shift operations. Such a procedure would require a very large number of computer based mimiced deals before a sufficient number of deals with 10-0 suit splits occurred. Thus, it seems advisable to attempt to construct another experiment which would require less computational effort to mimic. A little (or maybe considerable) thought will show that an equivalent experiment is to deal 26 cards to 2 hands with the restriction that there are only 3 cards of one suit, say spades, among the 26 cards. There may be 13 hearts and 9 clubs and 1 diamond or some other distribution of cards in the remaining three suits; the important point is that there are only 3 spades among the 26 cards. Now the computer mimicing of this deal requires far less computer effort. The RNG is used to produce an R, the R is compared to 3/26. If $R \leq 3/26$, we assign a spade to the first hand, and proceed to "deal the next card" to the second hand. If a spade was dealt to the first hand, the proportion of spades in the remaining 25 cards is now 2/25 and the next R must be compared to 2/25 to determine if a spade is to be dealt to the second hand. The procedure is now evident. To avoid unnecessary dealing of cards a test should be made to ascertain after the third card is dealt, if 3 spades have yet been dealt. If they have, the deal should be terminated and a new deal begun. A further shortening of the process can be attained by noting if there exists one spade in both hands. If this is the case, then there cannot possibly be three spades dealt to one hand so the deal should again be stopped and a new deal begun.

Thus it is seen that with some thought it is possible to construct experiments whose mimicing requires considerably less computer power than would other experiments. In this sense our computer estimation of probabilities is somewhat of an art, like programming. We have dwelt on this seemingly simple problem at some length just to illustrate some of the consideration necessary to effectively carry out a computer based mimicing.

We again remind the student that our purpose is to illustrate the great potential of the computer as an aid in quantitative analysis. The very act of constructing the experiment to be mimiced frequently contributes to a deeper understanding of the phenomena in question. It is hoped that the techniques of solution outlined above for these simple examples will be of assistance in the estimation of probabilities for more complicated problems. Sophisticated analysis of complex problems will require a good knowledge of probability theory and ingenuity with a computer.

Seventh Problem

In a certain large forest there are twice as many larch trees as there are white pine and three times as many spruce as white pine. It is known that 40% of the larch, 25% of the white pine and 33 1/3% of the spruce are suffering a worm infestation. A tree is selected at random and is found to be infected. What is the probability that the selected tree is a white pine? This question is equivalent to asking, "What proportion of the total number of infected trees is the number of infected white pine trees?". The hypothetical 'real' experiment consists in selecting a tree at random, determining which species it is, and then determining if it is infected. The experiment will be repeated a large number N, of times and a tally kept of the number of experiments in which a selected tree is infected and of the number of experiments in which a selected white pine tree is infected. In order to mimic the selection of the species of the tree it is first necessary to partition the unit interval into subintervals whose lengths are in the same ratios as the relative sizes of the corresponding species. Thus, since the unit interval corresponds to the entire forest, it is evident that 3/6 or 1/2 of the unit interval must correspond to spruce trees, 2/6 or 1/3 of the unit interval must correspond to larch and the remaining 1/6 of the interval should be assigned to white pine. The unit interval is thus partitioned as



Unit Interval Partition

The random selection of a tree from this distribution is mimiced by generating a random number $R1$ and comparing it to the fraction 1/2. If $R1 \leq 1/2$, a spruce tree is selected, if $R1$ is such that $1/2 < R1 \leq 5/6$ a larch tree is selected and if $R1 > 5/6$ a white pine is selected. For example, suppose $R1 = 0.638$; this means that a larch is selected. The health of the larch tree is determined by generating another random number $R2$ and comparing it to the fraction 1/4. If $R2 \leq 1/4$, the tree is said to be infected and a tally is made. If $R2 > 1/4$, the tree is said to be healthy and another tree selected from the forest. The process for arbitrarily selecting the species and then determining the state of health of the tree should now be apparent. The ratio of the number of infected white pine to the total number of infected trees is the desired probability.

4.7

The Eighth Problem

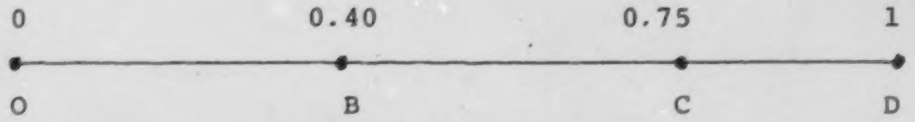
A large number of mice are treated for cancer by radiation therapy. It is noted that 40% improve, 35% remain unchanged and 25% actually become worse. Six mice are selected at random. What are the probabilities that:

(a) All six of the mice have improved?

(b) Two of the mice have improved, three remain unchanged and one has become worse?

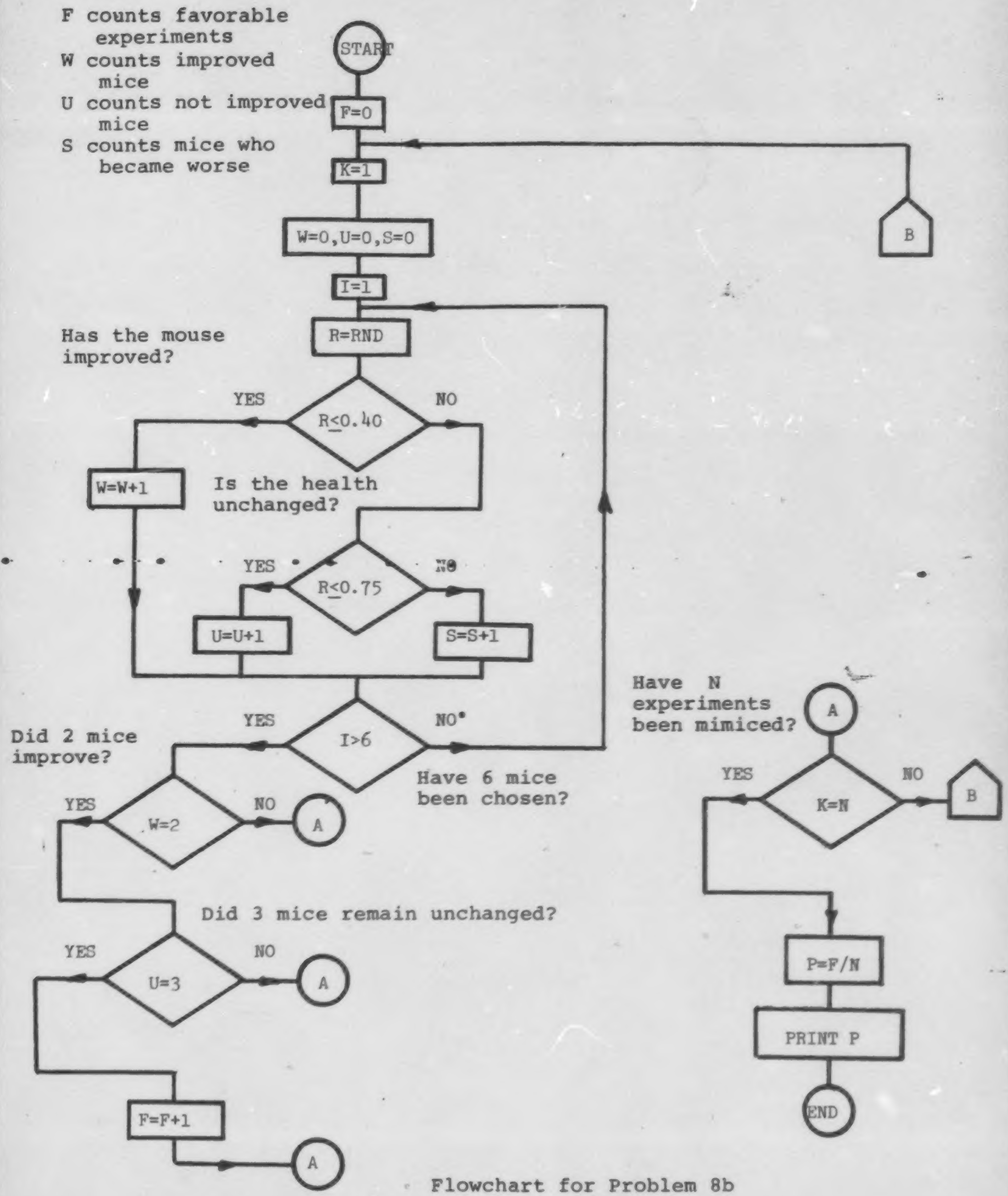
(c) One has improved, one remains unchanged and four have become worse?

The experiment to be performed is the random selection of six mice from those that have been treated for cancer radiation and the determination of the number of mice so selected whose health has improved, the number whose health has not changed, and the number whose health has actually become worse. In order to mimic the experiment, it is necessary to establish a method for randomly selecting the mice. This is easily done by recalling the procedure used to select an individual in problems 3, 5 and 7. Thus, the unit interval is partitioned into intervals of lengths 0.4, 0.35 and 0.25 respectively as shown below



UNIT INTERVAL PARTITION

In accordance with the data of the problem, the segment OA is assigned to those mice who have improved, the segment AB to those who showed no change and the segment BC to those who actually became worse. Using the RNG, a random number R is produced and the status of the health of the mouse is determined by noting which segment contains R. This process is repeated six times and each time the status of the mouse so chosen is noted and the appropriate sub tallies are kept. The experiment, which consists in mimicing of the selection of the six mice, the determination of the health of each mouse, and the appropriat tallying of the three states of health is repeated N times. For part (a), a tally is kept of the number of times all six mice have shown improvement. For part (b), tallies are kept of the number of times that two mice have improved, three mice have remained unchanged and one mouse actually became worse. Similar tallies are kept for part (c). The flowchart shown in figure 8.6 depicts the organization of the computer program to obtain an estimate to part (b).


F counts favorable experiments
W counts improved mice
U counts not improved mice
S counts mice who became worse
START
F=0
K=1
W=0,U=0,S=0
I=1
R=RND
Has the mouse improved?
YES
NO
R≤0.40
W=W+1
Is the health unchanged?
YES
NO
R≤0.75
U=U+1
S=S+1
YES
NO
I>6
Have 6 mice been chosen?
Did 2 mice improve?
YES
NO
W=2
A
Did 3 mice remain unchanged?
YES
NO
U=3
A
F=F+1
A
B
Have N experiments been mimiced?
A
YES
NO
K=N
B
P=F/N
PRINT P
END

Flowchart for Problem 8b

Fig. 8.6

Ninth Problem

In a northwestern community there is considerable discussion concerning the advisability of spraying forests with DDT to attempt to halt a spruce budworm infestation. The officials have decided to poll each of the community members of which there are 1750. The results of the poll reveal that 900 are in favor of using DDT, 500 are opposed to its use and the remaining 350 have no opinion. Nine people from the community are chosen at random. What is the probability that:

(a) All nine are in favor of spraying?

(b) Five are in favor of spraying and four are against it?

(c) Three are in favor, two are opposed and four have no opinion?

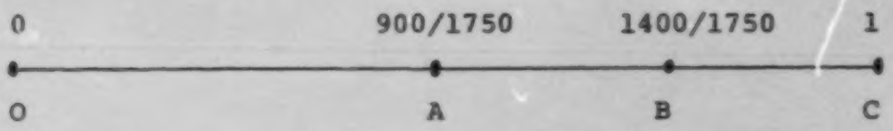
This is a problem with a finite number of individuals (elements or events) to choose from and so it is called a probability problem with a finite sample space. Data to calculate the estimates to all three parts of the problem can be obtained by repeated performance of the following experiment. Nine individuals in succession are selecte from the community and their opinion concerning the advisability or non-advisability of spraying with DDT is determined. For each such experiment a tally is made of the number in favor, F, the number opposed, A, and the number with no opinion, D. The individuals are replaced after each experiment and the entire experiment repeated N times. As each experiment is repeated the following tallies are kept:

(1) The number of experiments for which F=9. This will provide the data to estimate the answer to part (a).

(2) The number of experiments for which F=5 and A=4. This will provide the data to estimate the answer to part (b).

(3) The number of experiments in which F=3, A=2 and D=4. This will provide the data to estimate the answer to part (c).

The computer mimicing of the experiment requires the repeated partitioning of the unit interval. The determination of the opinion of the first individual to be chosen at random requires the initial partitioning of the unit interval into three subintervals of lengths 900/1750, 500/1750 and 350/1750 respectively. Pictorially this subdivision is represented on the unit interval as



UNIT INTERVAL PARTITION

The segment OA represents the people in favor of spraying, the segment AB represents the people opposed, and the segment BC represents those individuals who have no opinion concerning the advisability of spraying with DDT. The determination of a random sampling of opinions of each of the 9 successively selected individuals is accomplished by altering the three proportions in accordance with the resultant opinion expressed by the last chosen individual. The next paragraph gives an example of this.

The opinion of the first individual selected is determined by using the RNG to obtain a random number R1 and then noting the location on the unit interval of R1 relative to the points A and B. Suppose that R1=0.612; then it is seen that R1 lies in the interval AB and so we have selected an individual who is opposed to the spraying of DDT. As a result of having selected such an individual the proportion of the numbers of individuals holding the various opinions has changed, albeit not very much. The proportions now are: 900/1749 in favor, 499/1749 against and 350/1749 with no opinion. Thus, the selection of the opinion of the second individual requires the altering of the proportions used to select the opinion of the first or previous individual. Point A now has the coordinate 900/17 point B has the coordinate 900/1749 + 349/1749 or 1249/1749, and point C remains the end-point. A second random number R2 is selected and compared to these proportions to determine the opinion of the second individual. Suppose the number is R2=0.915, then the second individual is said to have no opinion and the new coordinates used to find the opinion of the third individual are 900/1748, 1399/1748 and 1. A third random number, R3, is selected and the process repeated for a total of nine times. The denominator used to determine the opinion of the ninth individual selected is 1742. A flowchart and computer program for the estimation of the probabilities of part c is shown in figure 8.7 and 8.8. The remarks inserted in the program should assist the student in following the work. On two runs of the program for 10,000 experiments probability estimates of 0.0247 and 0.0214 were obtained. Your author got careless and let the program run for 210,000 experiments and obtained the estimate of 0.022319. The actual probability is 0.0223423. See appendix A.

C - counts experiments
K - counts opinions in experiments

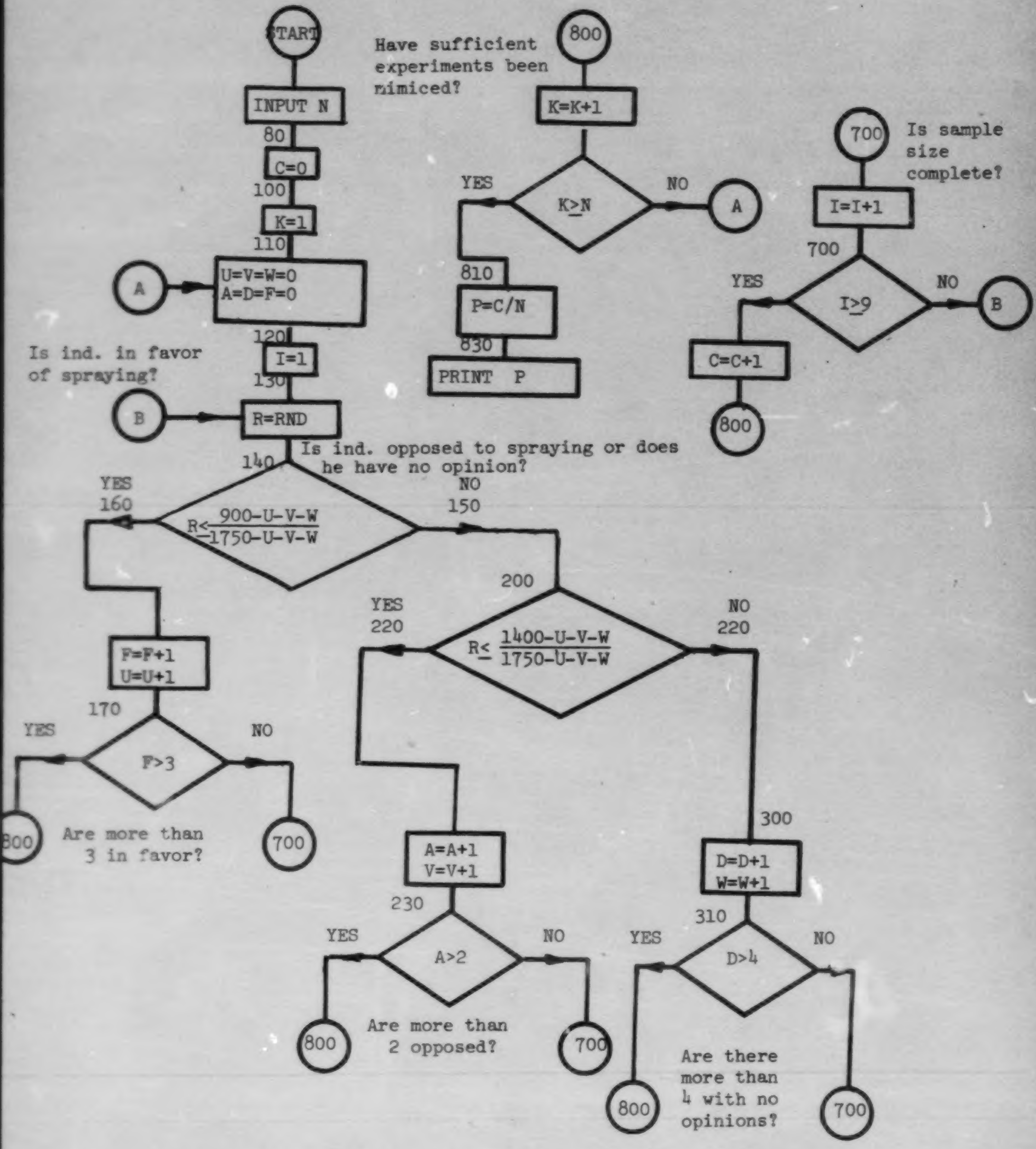

START
INPUT N
80
C=0
100
K=1
110
A
U=V=W=0
A=D=F=0
120
I=1
130
Is ind. in favor of spraying?
B
R=RND
140
Is ind. opposed to spraying or does he have no opinion?
YES
160
R≤ 900-U-V-W / 1750-U-V-W
NO
150
F=F+1
U=U+1
170
YES
F>3
NO
800
Are more than 3 in favor?
700
200
YES
220
R≤ 1400-U-V-W / 1750-U-V-W
NO
220
A=A+1
V=V+1
230
YES
A>2
NO
800
Are more than 2 opposed?
700
300
D=D+1
W=W+1
310
YES
D>4
NO
800
Are there more than 4 with no opinions?
700
Have sufficient experiments been mimiced?
800
K=K+1
YES
K≥N
NO
A
810
P=C/N
830
PRINT P
700
Is sample size complete?
I=I+1
700
YES
I≥9
NO
B
C=C+1
800

Fig. 8.7
Flowchart for Problem #9

```
1 REM THIS IS PROBLEM 9C
2 REM
3 REM
10 PRINT "TYPE IN THE NUMBER OF EXPERIMENTS, N"
15 INPUT N
16 PRINT
80 LET C=0
90 RANDOMIZE
94 REM
95 REM       LINES 110-120 INITIALIZE THE OPINION COUNTERS
96 REM              AT THE BEGINNING OF EACH EXPERIMENT
97 REM
100 FOR K=1 TO N
110 LET U=0
112 LET V=0
114 LET W=0
116 LET A=0
118 LET F=0
120 LET D=0
125 FOR I=1 TO 9
130 LET R=RND
134 REM
135 REM  INSTR. 140 DETERMINES IF IN FAVOR OF SPRAYING
136 REM
140 IF R<=(900-U)/(1750-U-V-W)GO TO 160
150 GO TO 200
160 LET F=F+1
162 LET U=U+1
164 REM
165 REM   INSTR. 170 DETERMINES IF MORE THAN 3 IN FAVOR
166 REM
170 IF F>3GO TO 800
180 GO TO 700
194 REM
195 REM    INSTR. 200 DETERMINES IF OPPOSED TO SPRAYING
196 REM
200 IF R<=(1400-U-V)/(1750-U-V-W)GO TO 220
210 GO TO 300
220 LET A=A+1
222 LET V=V+1
224 REM
225 REM   INSTR. 230 DETERMINES IF MORE THAN TWO ARE OPPOSED
226 REM
230 IF A>2GO TO 800
240 GO TO 700
300 LET D=D+1
302 LET W=W+1
304 REM
305 REM   INSTR. 310 DETERMINES IF MORE THAN 4 HAVE NO OPINION
309 REM
310 IF D>4GO TO 800
700 NEXT I
710 LET C=C+1
800 NEXT K
810 LET P=C/N
820 PRINT "PROB.  3 IN FAVOR, 2 OPPOSED & 4 WITH NO OPINION IS"
830 PRINT P
900 END
```

Fig. 8.8

The mechanism for changing the proportions in accord with the selection of an individual with a specified opinion is accomplished with the aid of counters. The counters are designated by the letters U, V and W respectively and they tally the respective numbers of individuals who are in favor of spraying with DDT, the numbers of individuals who are opposed to such spraying and the number who have no opinion about the spraying. These counters are used to properly alter the proportions in accord with the numbers of individuals who have previously been randomly selected and polled during the experiment. The incrementing of the counters is shown in figure 8.7 above the respective numbers 170, 230 and 310. The large diamond shaped decision figures contain the calculation necessary to establish the proper proportions. The term U+V+W represents the total number of individuals so far selected in the experiment. The expression

(900-U)/(1750-U-V-W),

which appears in the diamond under the number 140, is the fraction of the remaining total population that is in favor of spraying with DDT after a total of U+V+W individuals have been polled of which U were in favor of spraying. This method of calculating the fraction of the remaining population which is in favor of spraying is extremely useful in a great many probability problems wherein the probabilities change after each increment. The reader will recall that this procedure was introduced in the second problem in this chapter and also used in the chapter on Genetics.

It is interesting to note that by simply holding the proportions constant and equal to 900/1750, 500/1750 and 350/1750 respectively for the selection of of the 9 individual's opinions, we have mimiced the selection of opinions from a "large" population in which it is given that 900/1750 of the population favor the use of DDT, 500/1750 of them oppose its use and the remaining 350/1750 have no opinion. In our problem, because of the large number of people, 1750, and the small number of individuals in the sample size, 9, the successive proportions generated are nearly constant and each is very close to the respective proportions for the large population. Thus, with minimal loss in accurac and with considerable savings in computational effort, we could have used the original proportions to randomly select all 9 of the individuals. This fact is used to advantage statisticians because it greatly simplifies the calculation of the probabilities involving repeated trials. See the further discussion of this problem in appendix A. If, however, the number of people in the community had been smaller, say 100, and the size of the sample larger, say 35, then the proportions used in determining the last 10 or so opinions would have been considerably different than the proportions used to select the first 3 or 4 individuals. Thus, it may be seen that by running the program with different initial proportions and different sample sizes that the effect of both population and sample size can be investigated. This is another example of the flexibility of general computer programs.

Discussion

The technique of obtaining estimates to probabilities by mimicing the random process on a computer is an easy technique to learn and to use. However, the method, like any method, is not a panacea and must be applied with some care. Some of the considerations that need to be recognized in using the mimicing method are:

1. The possibility of excessive computational time required to mimic the required number of experiments.
2. The difficulty of assessing the correctness of the results. This is due to the fact that the mimicing process produces approximate results and, moreover, the results are not replicable because of the very randomness of the process.
3. The rapid increase in the computational effort to obtain more accurate answers. This consideration has been discussed earlier in the chapter.
4. The very heavy reliance on the random number generator subroutine. Because this subroutine is utilized so frequently, any inaccuracies it posseses, even if they are small, will certainly manifest themselves. We have not discussed the effect of such inaccuracies or inconsistencies in random number generator subroutines because our purpose is merely to illustrate an approach. This topic is discussed in texts on simulation.

Despite these difficulties, the method is useful and does assist in providing an intuitive understanding of random processes. In addition, the very fact that the results are not replicable, serves to confirm the expected lack of agreement when comparing empirically determined results of a random process with numerical results obtained from a theoretical analysis. By varying the number of experiments in these programs, the student can obtain an intuitive feel for the inherent variation in a random process thus acquiring some estimation of the possible variation in the expected difference of the results.

The mimicing method does permit the approximate analysis of random processes, which if analyzed in the classical way, would be intractable or at the very least would require computer based mimicing of part of the problem. The method frequently is useful when empirical data is used in conjunction with the theoretical formulation.

A final reason for presenting this method is the fact that the computer based simulation of stochastic processes is accomplished by mimicing a sequence of random events. Thus, the understanding of stochastic processes is facilitated by a familiarity with the mimicing method of estimating probabilities.

APPENDIX A

The Relation to Classical Probability Theory

The student who is familiar with the classical method for calculating simple probabilities should have noted the very close relationship between this method and the computer based method of estimating probabilities. Our computer based method of estimating the probabilities required three steps. They were:

(a) The visualization or construction of a hypothetical experiment which, if actually carried out a large number of times, would yield the data permitting an estimate of the probabilities.

(b) The construction of a computer program which, if run on a computer, would mimic the experiment.

(c) The actual running of the program a sufficiently large number of times in order to obtain the needed data.

It is the purpose of this appendix to illustrate how the analysis of these steps frequently can suggest a method for actually calculating the probabilities. The methods will require the repeated application of the notion of proportion and usually will involve only simple arithmetic. Our procedure will be to discuss each of the preceding examples and thus it will be assumed that the student is familiar with the preceding work. In particular, it will be assumed that the reader completely understands the statement of each problem, since in the interests of saving space, the problems will not be restated. We will not introduce such notions as conditional probability, Bayes' Theorem, etc. and then show how these notions can be utilized to obtain the solutions to certain classes of probability problems. Rather our approach will be to illustrate, by example, a general method of approach which should permit the student to solve many kinds of simple probability problems.

The elementary example of calculating the probability of throwing an ace on the single throw of a die has already been discussed on page 8.7. In that example it was noted that the interval (0,1/6) was 1/6 of the unit interval and this proportion was indeed the desired probability. A discussion of the second problem will be given later in this appendix. In the third problem, page 8.16, it will be recalled that in order to mimic the experiment it was necessary to subdivide the unit interval in sections whose lengths were equal to the respective proportions of the numbers of mice having the different characteristics. The pictorial representation of this subdivision is given on page 8.16. With the aid of the various lengths of the subdivisions it is quite apparent how the probabilities called for in this problem should be calculated. For part (a) the segment AB designates the proportion of mice with both afflictions. Since this segment is 0.1 of a unit, this means that 0.1 of the mice population have both afflictions. Thus, 1 out of 10 mice have both afflictions and according to the classical theory of probability this is the desired probability. For part (b) the section CD corresponds to the portion of the unit interval having no afflictions and so the proportion of the population having no afflictions is 35/100 which is the probability of a mouse having neither affliction as calculated in the classical manner. It is equally easy to answer the following problem. Suppose a mouse is selected at random from this population and that it has discolored eyes. What is the probability that it also has a short tail? Again, thinking of probability as a proportion this question is equivalent to asking what proportion of the population having discolored eyes is the population having short tails, i.e. what proportion of the segment AC is the segment AB? The answer is 0.1/0.25 or 0.4 and hence the desired probability is 0.4 or 4 out of 10.

The fourth problem, page 8.18, concerning the flies in the two cages necessitated the comparison of a random number R1 with the proportion 1/3 to determine if the randomly selected fly came from cage number one. Thus, 1/3 of the chosen flies would come from cage one and 2/3 would come from the second cage. The health of a randomly selected fly from the second cage was determined by comparing a second random number R2 with the proportion 1/4 because 1/4 of the flies in the second cage were ill. Thus, 1/4 of 2/3, or 1/6 of the original number of flies came from the second cage and were ill. Because we are assuming it is equally probable that any fly is selected from the original population, the proportion 1/6 is also the desired probability of selecting a sick fly from the second cage.

To answer part (b) of the fourth problem it is necessary to determine what is the proportion of flies that have a wing mutation but are not ill. This proportion will be determined by examining how the computer generated estimate of the probability was accomplished. Since the fly has a wing mutation it must have come from the second cage. Thus, it must have come from 2/3 of the original population. Now since R4 was compared to 3/10 to determine whether or not the fly had a wing mutation it is evident that 3/10 of 2/3, or 1/5 of the original population came from the second cage and had a wing mutation. R5 was compared to 1/4 to determine if a fly from the second cage was ill; thus 3/4 of the flies in the second cage were not ill. This means that 3/4 of 1/5, or 3/20 of the original population had a wing mutation and was not ill. Thus, the probability of selecting such a fly is 3/20.

The obtaining of the answers to the fifth problem, page 8.21, is not as straightforward as the obtaining of the answers to the previous four problems. Our discussion will be rather lengthy because it introduces a useful technique for obtaining probabilities. This technique is based upon the notion of calculating the number of favorable experiments in terms of the total number of experiments. These results are then used to calculate the desired proportions and these proportions are the probabilities. We will consider the parts of the problem separately.

The answer to part (a) requires the determination of the number of individuals from each city who have only measles since the total number of such individuals divided by the total population is then the required probability. To determine the number of such individuals in each city we follow the computer based experiment which first determines from which city the individual is selected and then determines whether or not the individual has only measles.

randomly selecting an individual from any one of the three cities, on the basis of the relative sizes of the three cities, implies that the number of individuals selected from a city is in direct proportion to the size of the city. Hence, if N individuals are selected, it follows that such a random selection will produce $6N/11$ individuals from city C1, $3N/11$ individuals from city C2 and $2N/11$ individuals from city C3. Now having determined how many individuals were selected from each city we again follow the computer program and note that of the individuals chosen from city C1 that $1/16$ of them have measles. Since $6N/11$ individuals come from city C1, it follows that, $6N/11 \times 1/16$, or $3N/88$ of them have measles, and because $1/8$ of the citizens of city C1 have influenza, $7/8$ of them do not have influenza. Thus, since the possibility of an individual having influenza is independent of whether or not he has a disease, it follows that $7/8$ of all of the individuals of city C1 do not have influenza. This implies that $7/8$ of $3N/88$, or $21N/704$ of the C1 citizens have measles but do not have influenza.

In a similar manner, of the $3N/16$ number of individuals selected from city C2, $3N/11 \times 3/10$, or $9N/110$ of these have measles. Because $9/10$ do not have influenza, $9/10$ of $9N/110$, or $81N/1100$ do not have influenza but do have measles. Finally, for city C3, $2N/11 \times 1/5$, or $2N/55$ have measles and $2N/55 \times 24/25$, or $48N/1375$ have measles but not influenza. Now the total number of such individuals from all three cities is $12177N/88000$ or $0.1384N$. Therefore, the probability of randomly selecting a person with only measles is $0.1384N/N$ or 0.1145. Your author apologizes for this rather long winded explanation. However, he wanted to emphasize that by reasoning in analogy with, and in the order of, the computer programs the answer can be obtained.

The answer to part (b) is obtained in a very similar manner. In contfast to part (a), it is necessarv to obtain the proportion of the number N, of people selected that have both measles and influenza. Of the 6N/11 people in C1, 1/8 of them or 3N/44 have influenza. Since the proportion of all citizens of C1 having measles is 1/16, it follows that 1/16 of 3N/44, or 3N/704 individuals from this city have both diseases. Analogously, 3N/11 x 1/10, or 3N/110 citizens from city C2 have influenza and 3N/110 x 3/10, or 9N/1100 have both diseases. Finally, 2N/11 x 1/2 or 2N/275 of the residents of C3 have influenza and 2N/275 x 1/5, or 2N/1375 have both diseases. Thus, the total number of people from all three cities having both diseases is 1223N/88000. Since the proportion of people from all three cities having both diseases is the total number of individuals having both diseases divided by the total number of residents, the proportion is (1223N/88000)/N or 0.0139. This is the required probability.

Part (d) of the problem is equivalent to determining the number of residents of city C2 having both diseases is what proportion of the total number of residents having both diseases. Since the number of residents of each city having both diseases has already been calculated above in obtaining the answer to part (b), the desired proportion is

$$(9N/1100) / (1223N/88000) \quad \text{or} \quad 720/1233 = 0.5887$$

This is the desired result.

The obtaining of the answer to part (c) is more difficult and requires a different method of attack because the experiment is different. Here the experiment consists of drawing 3 individuals, one from each city, and it is this experiment that is repeated N times. This results in the selection of 3N individuals and it is the determination of what proportion of these 3N individuals is the number of individuals selected such that they have only one disease among the three simultaneously selected. Thus, the number of such individuals must be determined. The student will note that, for each experiment, one citizen is selected from each city and so the relative sizes of the cities is not a consideration in the calculation. The number of individuals from the first city that have only one of the diseases is obtained by adding the number of individuals who have measles but not influenza to the number who have influenze but not measles. Thus for city C1, which is assumed to have N people, N/16 citizens have measles and 7/8 x N/16, or 7N/128 have measles but not influenza. Similarly, N/8 of the people have influenza and 15/16 of N/8, or 15N/128 have influenza but not measles. Thus the number of people from city C1 having only one of the diseases is 11N/64. A similar calculation shows that there are 7N/100 people in city C2 who have influenza but not measles and 27N/100 who have measles but not influenza for a total of 17N/50 people who have only one of the diseases. Finally, N/25 x 4/5 plus N/5 x 24/25, or 28N/125 of the people in city C3 have only one of the diseases.

It is now necessary to determine the number of individuals chosen from city C1 who have only one of the diseases and which corresponds to a selection of a pair of individuals from the other two cities having no diseases. A similar determination must be made for cities C2 and C3. This determination is possible because we are assuming that the choice of a diseased or healthy individual from any city does not depend upon the health of the residents of the other cities. This means for example that if a healthy individual is chosen from the first city, this fact has no effect on the status of the health of the individuals chosen from the other two cities. Thus, if for example, 3/4 of the individuals in city C2 are healthy then because the individuals are picked simultaneously from each city, 3/4 of the citizens selected from cities C1 and C3, regardless of their health, will be accompanied by the selection of a healthy individual from city C2. Hence, if 400 sick persons are selected from city C1 then 300 healthy persons are simultaneously selected from city C2. Thus, from both cities 300 individuals are selected who are ill. If, in addition 3/5 of the individuals chosen from city C3 were healthy, then 3/5 of 300 or 180 individuals selected from all three cities would be ill and these 180 would all come from the first city. It is seen therefore that out of N individuals selected the calculation of the number of individuals selected from a given city who have only one diseas and who are accompanied by the choice of a healthy individual from each of the two remaining cities requires the determination of the proportio of healthy individuals in each of the cities.

The proportion of healthy individuals selected from city C1 is obtained by first calculating the number of healthy members of C1. The number of individuals in the first city who have either disease or both diseases is N/16 + N/8, or 3N/16 and hence the number of healthy individuals is 13N/16. Thus the proportion of healthy individuals in the first city is 13/16. A similar calculation for city C2 reveals that the proportion of healthy individuals is 3/5 and for city C3 the proportion is 19/25. Hence, when choosing an individual from each city the probability of choosing a healthy individual from the first city is 13/16, from the second city 3/5, and from the third city 19/25.

Since the number of people from the first city that have only one of the diseases is 11N/64, 3/5 of these, or 33N/320 will correspond to the simultaneous selection of all healthy individuals from city C2 and in turn 19/25 of these, or 627N/88000 will correspond to the simultaneous selection of no sick individual from city C3. Hence, of the total number, 3N, of individuals selected from all three cities, the number of individuals having only one of the diseases and coming from the first city is 627N/8000. Similarly, the number of individuals selected from the second city having only one disease is 17N/50 and 13/16 of these, or 221N/800 correspond to the simultaneous selection of a healthy individual from the first city. Then, because 19/25 of the individuals chosen from the third city are healthy, 19/25 x 221N/800 or 4199N/20000 individuals selected from the second city are simultaneously selected with healthy individuals from each of the other two cities. Finally an analogous calculation shows that 273N/2500 individuals selected from the third city are simultaneously picked with healthy individuals from the first and second cities.

Hence, the total number of individuals selected such that they only have one disease among the three simultaneously selected is

$$\frac{627N}{8000} + \frac{4199N}{20000} + \frac{273N}{2500} \quad \text{or} \quad 0.397525N$$

The proportion so chosen is $0.397525N/3N=.1325$ and this is the requir probability. Whew!!!

The method of solution of this problem was based upon calculating the numbers of individuals having various characteristics and then forming the desired proportions or ratios. The numbers were expressed as fractions of N, where N was the number of times the experiment was repeated. This notion of expressing the numbers of individuals or events in terms of a fraction of the number of experiments is a very fruitful one. In fact, almost all of the problems could have been solved using this idea. As an example we "redo" the third problem using this notion of expressing the number of occurrences of the various events in terms of the number N of experiments.

From the data of the problem, $N/2$ of the mice have short tails, $N/4$ have discolored eyes and 0.40 of $N/4$ or $N/10$ of the mice have both afflictions. Thus, $N/4 - N/10$ or $3N/20$ have only discolored eyes and $N/2 - N/10$ or $2N/5$ have only short tails. To answer part (a) we note that the proportion of mice having both afflictions is $(N/10)/N$ or $1/10$ which agrees with the answer obtained previously. The answer to part (b) requires the determination of the number of mice having no afflictions. This number is given by subtracting the number of mice having only short tails, the number having only discolored eyes and the number having both afflictions from the total number, N. Thus, $N - 2N/5 - N/20 - N/10=7N/20$ mice are healthy. The probability is then given by the ratio $(7N/20)/N$ or $7/20$ which again agrees with the previous result. The student should work out on his own the fourth problem using the "number of experiments" method. It is a very useful and quite intuitive technique and for these reasons we shall frequently use it.

To answer the sixth problem, page 8.31, we note that part (a) is similar to part (c) of the fifth problem. The experiment consists in selecting a tree from each forest and a tally is to be made whenever only one of the three trees selected is healthy. Now the problem is to calculate the proportion of the number of experiments that resulted in the selection of exactly one infected tree of the three trees chosen. Because the experiment is done a large number of times part of the desired tallies will correspond to cases wherein the infected tree came from the first forest and simultaneously only healthy trees came from the other two forests. Similarly, part of the tallies will correspond to infected trees being chosen from the second forest with the other forests having had healthy trees selected from them. An analogous remark may be made about the infected trees from the third forest. It is necessary to calculate the number of trees corresponding to each of the above cases in order to arrive at a total.

The number of infected trees selected from the first forest is N/3. Simultaneous with the selection of N/3 infected trees from the first forest was the selection of a number of healthy and unhealthy trees from the second and third forests. Because 3/4 of the trees in the second forest are healthy, 3/4 of the selections of any tree from the first forest will be accompanied by the selection of a healthy tree from the second forest. It is assumed that there is no preference in the selection of the trees and this assumption holds throughout the problem. Thus 3/4 of N/3 or N/4 of the trees selected from the first forest were infected and accompanied by the choice of a healthy tree from the second forest. Similarly, because

$4/5$ of the third forest is made up of healthy trees, $4/5$ of $N/4$ or $N/5$ trees selected from the first forest are infected and were selected simultaneously with healthy trees from the other two forests. The procedure for calculating the number of cases wherein exactly one tree was infected and the tree came from the second forest is the same. Thus, of the N trees selected from the second forest, $N/4$ of them are infected. $2/3$ of $N/4$ or $N/6$ of the selected trees are infected and are accompanied by a choice of healthy trees from the first forest. Finally, the number of infected trees chosen from the second forest and corresponding to the choice of no infected trees from the first or third forest is $4/5 \times N/6$ or $2N/15$. Similarly the number of infected trees chosen from the third forest along with no infected trees from the other two forests is $N/5 \times 2/3 \times 3/4$ of $N/10$. Therefore, the total number of experiments in which exactly one infected tree was selected is

$$N/5 + 2N/15 + N/10 \qquad \text{or} \qquad 13N/30.$$

The proportion, and hence the probability, is $13/30$.

The problem posed in part (b) is equivalent to the determination of what proportion is the number of times exactly one infected tree is selected and that tree comes from the second forest to the total number of times in which exactly one of the selected trees was infected. From part (a), the number of experiments resulting in the selection of only one infected tree and that tree came from the second forest is $2N/15$. In addition, the number of experiments which correspond to exactly one infected tree regardless of which forest it came from, was $13N/30$. Therefore, the desired probability is

$$\left(\frac{2N}{15}\right) \Big/ \left(\frac{13N}{30}\right) \quad \text{or} \quad 4/13.$$

The calculation of the probability required in the seventh problem, page 8.36, is straightforward. The student should reread the section describing the computer mimicing of the experiment. We begin our calculation by letting N be the number of experiments and calculate the number of infected trees of each species in terms of N. Now in accord with the partitioning of the unit interval, N/2 of the trees selected will be spruce, N/3 will be larch and N/6 will be white pine. Of the white pine, 1/4 are infected and therefore, the number of infected white pine trees is N/24. A similar calculation shows that the number of infected larch is 2N/15 and the number of infected spruce is N/6. Thus, the probability is

$$\frac{N/6}{\frac{N}{24} + \frac{2N}{15} + \frac{N}{6}} \qquad \text{or} \qquad 20/41$$

We now consider the calculation of the probabilities asked for in the second problem, page 8.11. Since the experiment of selecting 5 cows is repeated N times, the answer to part (a) is obtained by calculating the number of experiments in which all five of the selected cows had not recovered. This calculation is done in the following way. Because the herd consists of 5 animals who have not recovered, 5N/20 is the number of experiments in which a sick cow may be selected on the first draw. For these selections, since a sick cow was selected, 4 out of the 19 remaining cows are ill. Consequently, 4/19 of the N second draws will result in a sick cow being chosen. This means that of the 5N/20 experiment which resulted in the initial selection of a sick cow, only 4/19 of these, or N/19 experiments will result in a sick cow being selected on both the first and second draw. For these cases, 3/18 of the experiments will result in the third draw yielding an ill cow. Thus, of the N/19 experiments resulting in the selection of 2 ill cows on the first 2 draws, only 3/18 of them, or N/114 experiments will result in the selection of a sick cow on each of the first three draws. The procedure for completing the calculation should now be evident. 2/17 of N/114, or N/969 experiments will result in the first four draws selecting an ill cow on each draw. Finally, 1/16 x N/969 or N/15504 experiments will result in recovered cows being selected on each of the five draws. Therefore, the probability that none of the 5 selected cows has recovered is 1/15504.

The student should have noticed that this answer was obtained by multiplying the successive proportions of sick animals, i.e.

$$5/20 \times 4/19 \times 3/18 \times 2/17 \times 1/16 = 1/15504.$$

This fact suggests how to calculate such probabilities for similar problems. If these proportions are interpreted as probabilities, the above calculation is an example of the product law in classical probability.

Part (c) is done in a very similar manner. The experiment is the same, but the initial and successive proportions are different. Our method of explanation will be to again work out the number of successful events for each successive draw of an animal. Out of N selections of 5 animals per each experiment, 15N/20 selections will consist in having chosen a recovered animal for the first of the 5 animals to be selected. This being the case, the proportion of remaining healthy animals is 14/19. Hence, 14/19 of 15N/20 or 21N/38 experiments wil result in the selection of 2 healthy animals on the first two draws. Using the hint described in the previous paragraph, the procedure for calculating the probability of picking three healthy cows on the remaining three draws should be evident. The probability is

15/20 x 14/19 x 13/18 x 12/17 x 11/16 or 351/5168.

Part (b) is more difficult and requires an examination of the computer based mimicing of the selection process. It will be recalled that the program shown estimates the probability that all 5 cows have recovered and that by changing line 180 to read

```
180 IF J = 3 GOTO 200
```

the estimation of the probability of drawing exactly 3 recovered animals may be obtained. Because only 3 recovered animals are to be chosen, an examination of how the computer could pick exactly three recovered cows, reveals that it could have accomplished this selection in a variety of ways. For example, a healthy cow could have been picked on the first, third and fourth draws and a sick cow picked on the second and fifth draws, or healthy cows could have been chosen on the second, fourth and fifth draws and cows still suffering from hoof and mouth disease picked on the first and third draws. Other orders of selection (combinations) are also possible which would result in the selection of exactly 3 healthy cows being picked out of the 5 animals drawn. Furthermore, each of these selections appear to be equally probable since there is no priori reason that the random selection procedure should prefer one of these combinations to another such combination of 5 selected cows. Thus, the calculation of the probability that exactly 3 recovered cows were chosen out of the 5 selected requires:

(a) The calculation and listing of all possible combinations of a sample of 5 animals containing exactly 3 recovered cows,

(b) The calculation of the number of experiments that will result in the selection of each of the combinations found in part (a), and

(c) Adding the results obtained in part (b), and dividing by N to obtain the desired probability.

We will next illustrate how each of these calculations may be performed and then we will show that the above set of tasks is equivalent to carrying out the following 4 calculations.

(1) The calculation of the number of ways in which 3 distinct objects may be selected from 5 objects.

(2) The calculation of the number of experiments corresponding to any one combination of 5 draws yielding exactly 3 recovered animals.

(3) The multiplication of the number of experiments obtained in step 2 by the number of ways obtained in step 1.

(4) Dividing the result of step 3 by N, the number of experiments.

The latter set of 4 calculations requires much less computational effort than th former set of 3 calculations. For those students who are familiar with probability, the procedure is equivalent to using the binomial distribution to calculate the probabilities for a finite sample space as is done in classical probability theory.

Now the listing of the combinations of 3 healthy animals from 5 animals can be accomplished with the aid of a computer program. In order to not interrupt the discussion of the method for calculating the probability, we will simply list the possible combinations. A description of a computer program which both lists and counts the combinations is given in Appendix B. Let R and S denote the selection of a recovered animal and a sick animal respectively; then the possible selections are: RRRSS, RRSRS, RRSSR, RSRRS, RSRSR, RSSRR, SRRRS, SRSRR and SSRRR. Hence, there are 10 possible selections of exactly three recovered animals out of 5 animals. In the computer based experiment, to generate the required tallies for problem 2, if the number of experiments, N, is very large, each of these combinations will be selected several times by the computer. This suggests that we calculate the number of experiments resulting in the choice of each of these combinations. Consider the choice RRSRS. At the risk of boring the student, we work this calculation out in complete detail so that the succeeding comments and work are easier to appreciate. From our previous work, the number of experiments resulting in the selection of an R on the first draw is 15N/20 and the number of experiments resulting in the selection of an R on each of the first two draws is 14/19 x 15N/20 or 21N/38 experiments. Since the proportion of sick cows in the remaining herd is 5/18, it is evident that 5/18 x 21N/38 or 35N/204 experiments resulted in the selection of a recovered cow on each of the first two draws and the selection of a sick cow on the third draw. Now, 13/17 of the remaining animals have recovered and so the number of experiments resulting in the selection of RRSR on the first four draws is 455N/3468. Finally, since 4/16 of the herd has not recovered, the number of experiments resulting in the choice RRSRS is 455N/13872.

A review of this calculation reveals that the result was obtained by multiplying the respective proportions 15/20, 14/19, 5/18, 13/17, and 4/16 and then multiplying the result by the number of experiments. Furthermore, the proportions were obtained by noting whether a recovered or a sick cow was selected and then subtracting one from the respective number of remaining healthy or sick cows. The product was divided by the number of animals remaining in the herd to give the respective successive proportions. The analogy between this process and line 140 of the program listed in figure 8.2 should be noted. This procedure for calculating the number of experiments corresponding to a particular combination of 5 cows wherein exactly 3 of them have recovered suggests that the number of experiments resulting in any other such combination, for example SRRSR, can be obtained by successively multiplying the respective proportions 5/20, 15/19, 14/18, 4/17 and 13/16 and then multiplying by the number of experiments. Thus, the number of experiments yielding the combination SRRSR is 5/20 x 15/19 x 4/17 x 13/16 x N or 455N/13872. This is the same as the previous result and should not be surprising since the previous result may be written as 15/20 x 14/19 5/18 x 13/17 x 4/16 x N and a comparison of the two products shows that the multiplicands in the numerators and denominators respectively are the same but appear in a different order. The student should veri for himself that the same result holds for the remaining eight combinations. In fact, the generalization of this result is clear and is ver useful. Because the number of experiments is the same for each choice and there are ten possible choices, the total number of experiments resulting in the selection of exactly 3 recovered cows out of the 5 selected is 10x455N/13872 or 2275N/6936 and hence, the desired probability is 2275N/6936.

We will apply this procedure to another similar problem in order to assist the student in fixing the procedure in his mind. Suppose that a grove of 50 spruce trees has been infested by a spruce bud worm infestation and that 20 of the trees had become infested. If seven trees are selected at random, what is the probability that exactly four had become infested? As before, it is first necessary to recognize that the experiment to be mimiced consists in the random selection of 7 trees from the grove and the attendant determination of the well being of each of the 7 trees. The experiment is to be simulated N times on the computer and the above method of calculating the probabilities consists in the determination of the number of experiments resulting in exactly four infested trees being among the seven trees selected in the sample. If it is known that there are four infested trees in the sample, there are many ways in which four infested trees can be chosen from the sample of seven trees. Now, the total number of experiments resulting in the selection of exactly four infested trees from the seven which are sampled is determined by the product of the number of ways in which four such trees can be selected from seven trees and the number of experiments resulting in just one of the combinations of four trees out of the seven. By actually listing the possible combinations it may be seen that there are 35 such possible ways. As an example, two of the combinations are HHSSHSS and SHSHSHS where H and S denote a healthy and an infested tree respectively. Since N experiments are performed, the number of experiments in which the combination HHSSHSS is achieved is

$$30/50 \times 29/49 \times 20/48 \times 19/47 \times 28/46 \times 18/45 \times 17/44 \times N$$

or 9367N/1664740. The corresponding expression for the number of experiments in which the combination SHSHSHS is selected is

$$20/50 \times 30/49 \times 19/48 \times 29/47 \times 18/46 \times 28/45 \times 17/44 \times N$$

or 9367N/1664740, and it is seen that the two expressions are equal in numerical value. The equality of the two results should not be surprising in view of the previous discussion. Since the total number of experiments resulting in a choice of exactly four infested trees out of the seven selected is N/35, the desired probability is then 35/N times the above expression or 9367N/47564. This completes our discussion of the second problem.

We now consider the solution of the eighth problem. The calculation of the number of experiments for each part of the problem is accomplished in a manner entirely analogous to that developed to calculate the similar quantities in the second problem. For part (a), since there is only one possible way of selecting six mice and since the proportion of mice that improved is 2/5, it follows that $(2/5)^6 xN$ is the number of mice that improved in N experiments. The proportion interpretation of probability then implies that the probability of choosing an improved mouse on each of the six draws is $(2/5)^6$ or 0.004096.

The answer to part (b) is obtained by following the four steps listed in the discussion of the second problem. The number of ways of selecting two improved mice, three mice which are unaffected and one mouse which has become worse from six such mice can be obtained by the use of the computer program described in Appendix B. This may be done by letting the digit 3 denote an improved mouse, the digit 2 an unaffected mouse and the digit 1 a mouse who became worse. The computer will then list the 60 combinations. The next step is to calculate the number of experiments corresponding to just one of the combinations of six mice. We will describe the calculation for six mice listed in the order 3, 3, 2, 2, 2, 1. The number of experiments resulting in the selection of an improved mouse on both the first and second draws is $(2/5)^2 xN$. Since the proportion of unchanged mice is now 7/20, the number of experiments resulting in the selection of improved mice on the first two draws and unchanged mice on each of the next three draws is

$(2/5)^2 \times (7/20)^3 \times N$. Finally, because 1/4 of the mice are actually in worse shape after the treatment, the number of experiments resulting in the required combination is $(2/5)^2 \times (2/20)^3 \times (1/4) \times N$. Since there are 60 such combinations the total number of experiments is $(2/5)^2 \times (7/20)^3 \times (1/4) \times 60 \times N$ and therefore, the required probability is 1029/10000.

Part (c) is done the same way. There are 30 ways of selecting one improved mouse, one unchanged mouse and four mice which have become worse from six such mice. By now the student should have recognized that the number of experiments corresponding to one specified combination of the 30 combinations is merely the product of the respective proportions corresponding to the state of health of the mice. Thus $2/5 \times 7/20 \times (1/4)^4 \times N$ is the number of experiments in which one specified combination is selected. The total number of experiments corresponding to all possible combinations is $2/5 \times 7/20 \times (1/4)^4 \times 30 \times N$ and hence the probability is 21/1280.

The calculation of the designated probabilities for the ninth problem requires a slight alteration of the procedure for calculating the number of experiments corresponding to a specific combination. This alteration is due to the fact that the population is finite and thus, as citizens are selected, the proportion of the remaining individuals holding a prescribed opinion changes. The situation is entirely analogous to that encountered in the second problem. Part (a) requires the calculation of the number of experiments resulting in samples all of whose citizens opposed the spraying. By recalling the method of calculation used in the second problem it is seen that the number of such experiments is

$$500/1750 \times 499/1749 \times 498/1748 \times 497/1747 \times 496/1746 \times 495/1745 \times 494/1744 \times 493/1743 \times 492/1742 \times N \quad \text{or} \quad 1.20473 \times 10^{-5} \times N.$$

Since there is only one way to select 9 such opinions from 9 people this number is also the total number of experiments in which all 9 people are opposed to spraying with DDT. The probability is then this number divided by N or 1.20473×10^{-5}.

For part (b), the number of possible combinations of 9 opinions of which 5 are opposed to spraying and 4 are in favor of spraying is 126. To calculate the number of experiments corresponding to just one of these combinations, we note that the number of experiments in which the first 5 draws in succession result in an opposing vote is

$$500/1750 \times 499/1749 \times 498/1748 \times 497/1747 \times 496/1746 \times N.$$

Hence, the number of experiments resulting in the specified combination for all 9 opinions is

$$500/1750 \times 499/1749 \times 498/1748 \times 497/1747 \times 496/1746 \times 900/1745 \times 899/1744 \times 898/1743 \times 897/1742 \times N \quad \text{or} \quad 1.32378 \times 10^{-5}N$$

Multiplying this number by 126 to obtain the number of experiments corresponding to all possible combinations and then dividing by N gives a probability of .00166796.

The number of possible combinations in part (c) is 1260 and the number of experiments corresponding to just one of these combinations is

$$900/1750 \times 899/1749 \times 898/1748 \times 500/1747 \times 499/1746 \times 450/1745 \times 349/1744 \times 348/1743 \times 347/1742 \times N \quad \text{or} \quad 1.77320 \times 10^{-5}N$$

The probability is obtained by multiplying this number by the number of combinations and dividing by the number of experiments to give 0.0223423.

Your author purposely chose this problem because it illustrates the computational difficulties than can arise even when an exact answer can be obtained. The numerical evaluation of the products of the fractions is time consuming and, when done on a computer, can lead to serious loss of accuracy since only limited precision arithmetic can be used. It is also the case that overflows can easily occur in the computer if the order of the arithmetic is not chosen properly. For example, suppose that the method of numerically evaluating the product is to first multiply all of the numerators, then multiply all of the denominators and then divide the former product by the latter product. As the student can note, each of these products could be quite large. In fact, if the sample had contained 80 individual's opinions, the products would be larger than 10^{100}. One method for minimizing the numerical difficulties associated with the evaluating of such products is to alternate the order of the operations of multiplication and division. Special programs have been developed to evaluate such products and your local computer center should be able to provide them.

The initial proportion of citizens in favor of using DDT is 900/1750 and the initial proportion opposed to the use of DDT is 500/1750 whereas the initial proportion having no opinion is 350/1750. Now, if these proportions were representative of a "large" population, that is a population in which it is assumed that the proportions do not change as individuals are selected, then the answer to part a would be $(500/1750)^9$ or 1.2688×10^{-5} which is very close to the previous answer. Thus, again we see that if the ratio of the total population to the sample population is greater than 10 or so that very little error is made if no correction is made to the proportions while doing the calculation.

Similarly, the answer to part (b) would be $(500/1750)^5 \times (900/1750)^4 \times 126$ or 0.0167822 and the answer to part (c) would be $(900/1750)^3 \times (500/1750)^2 \times (350/1750)^4 \times 1260$ or 0.0223856.

APPENDIX B

In the previous appendix it was stated that a listing of all possible ways of selecting 3 recovered cows from 5 cows could be accomplished with the aid of a computer. This appendix describes the development of the computer program for obtaining such a listing. Even though such a listing was not needed to calculate the probability (only the number of such ways was needed), the program may be of interest in its own right. The program to be developed will list all of the possible distinct combinations of n objects and, if some of the objects are identical, the listed combinations will be distinct. It is the ability to list only distinct combinations which makes the program useful since the listing of all possible combinations, regardless of whether or not they are distinct, is not very helpful.

If a recovered cow is associated with the digit 2 and a cow who has not yet recovered associated with the digit 1, it is evident that the determination of the number of possible ways of selecting 3 recovered cows from a group of 5 cows is equivalent to determining the numer of possible orderings of the numbers 2, 2, 2, 1 and 1. There are ten such orderings and they are: 2,2,2,1,1; 2,2,1,2,1; 2,2,1,1,2; 2,1,2,2,1; 2,1,2,1,2; 2,1,1,2,2; 1,2,2,2,1; 1,2,2,1,2; 1,2,1,2,2; and 1,1,2,2,2. The order of the digits describes the order in which the cows were chosen. For example, the order 2,1,2,1,2 corresponds to first choosing a recovered cow and then alternately choosing a cow which has not yet recovered followed by a recovered cow, etc. This suggests the development of a program which receives as input several numbers and then as output lists them in all possible distinct orders. For example, if it is desired to list all possible ways of selecting 3 recovered cows from a group of 5 cows, the input would be the five digits 2,2,2,1 and 1 and output would be the 10 five digit numbers as listed above. As a further example, suppose it is desired to know the number and ways in which it is possible to select 2 brown horses, 2 red horses and 1 black horse from a group of 5 such horses. By identifying the digit 3 with a brown horse, the digit 2 with a red horse and the digit 1 with a black horse, the input to the program would be the set of numbers 3,3,2,2, and 1. The possible orderings are (for brevity the commas are omitted): 33221, 33212, 33122, 31322, 31232, 31223, 13322, 13232, 13223, 12323 and 122233 and the number of such selections is 12. Each set of numbers prescribe an order in which the horses are to be chosen.

Each of the preceding sequences of numbers were listed in decreasing order and it is the property of decreasing order that shall be exploited in the development of the computer program. It is helpful to note that because the numbers of the listing are to appear in decreasing order that any number in the list is smaller than all those preceding it and larger than all those numbers following it in the list. This fact provides the key to the development of the program because it ensures that all of the numbers will be distinct, i.e. there will be no repetitions, even if the original number has repeating digits. With these remarks in mind we proceed to discuss the following problem. Suppose the first K numbers in the list have been constructed, how then is the next number to be constructed? Such a construction requires a procedure for insuring that the largest of all possible smaller numbers is the next number selected. To this end we examine a numerical example and see if we can "divine out" a method.

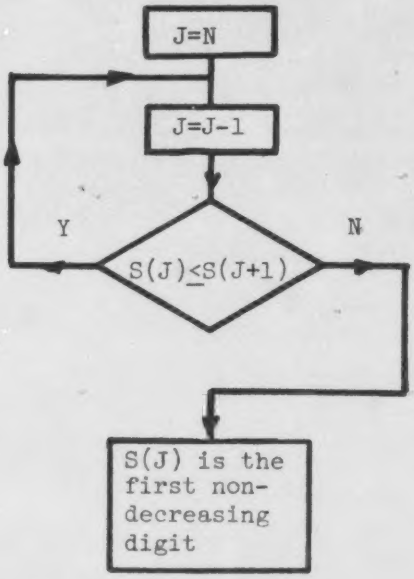
First of all, we adopt the convention that the numbers for which we are going to find all possible listings must be entered in such a way that the number formed by them is the largest possible such number. Thus, the initial arrangement of the digits must be such that they do not increase as they are read from left to right. For example, if the digits 5, 9, 1, 3, 8 are to be so listed, they must be entered in the order 9, 8, 5, 3, 1 and if the digits 3,4,3,5,4,6,5 are to be so listed, they must be initially entered in the order 6, 5, 5, 4, 4, 3, 3.

For our numerical example, we consider the digits 5,8,1,3,9 and for the sake of convenience we will omit the commas when writing them. The first several entries in the output of the computer program should appear as

```
98531
98513
98351
98315
98153
98135
95831
95813
95381
95318
95183
95138
93851

 etc.
```

There are 120 such numbers and the student should extend the list up to 50 or so. Reading from the bottom up, the last few numbers in the list are 13589, 13598, 13859, 13895, 13958, 13985, 15389, 15398, etc. The student should study this list and attempt to make up the rule for determining the subsequent entries in the table.

Now we recall that if any sequence of digits is in non-decreasing order when read from left to right then the number formed by these digits is the smallest number that can be formed by them. This fact, when applied to the number 95138 shows that any interchange in the last three digits 1, 3, or 8 would only serve to increase the numerical value of the number 95138. The digit in the second place, 5, is the first digit not in decreasing order when reading from right to left. It is the first digit out of non-decreasing order and it is this very property that indicates that the digit 5, and all the digits to the right of it must be interchanged to insure the obtaining of the next smallest number in the listing. The determination of how these digits must be interchanged is fascilitated by the introduction of the following notation. Let $S(J)$ denote the digit J places from the left. Thus, for the number 95138 we have $S(1)=9$, $S(2)=5$, $S(3)=1$, $S(4)=3$ and $S(5)=8$. In this example the sequence $S(3)$, $S(4)$ and $S(5)$ was non-decreasing and $S(2)$ was the first digit out of non-decreasing order. With the aid of this notation, a flowchart of the procedure for determining the first digit, counting from the right, that is in non-decreasing order is:

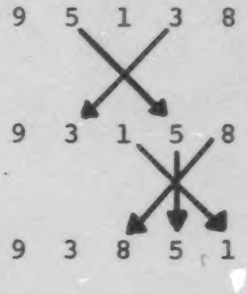


Further study of the sequence reveals that the out of order digit is always interchanged with the next smaller number to the right of the out of order number. Morever, there will always be such a number unless the end of the listing has been reached. If this interchange is made, the number 95138 is transformed to 93158 which is not the next smaller number. However, it is evident that by reversing the order of the last three digits to give 93851 that the next smaller number is obtained. Thus a suggested procedure for obtaining the next smaller number consists of the four steps in the order in which they are stated:

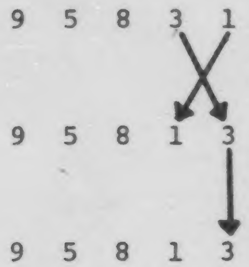
1. Reading the number from right to left and determining the first digit which is increasing and denoting this digit by S(L).
2. Reading from left to right beginning with the digit S(L) and determining the next smaller digit to S(L). Denote this digit by S(R).
3. Interchanging S(L) and S(R).
4. Reversing the order of all of the digits to the right of the L^{th} digit.

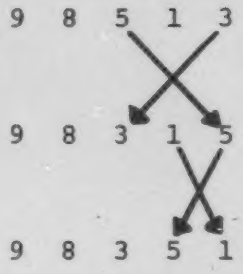
It should be noted that these steps have been devised from an examination of a listing of a number all of whose digits are distinct. The steps may have to be modified if the number is something like 95533. We are considering the easiest case first. The student should recall that this is in keeping with our philosophy of first devising a program to solve the easy problem on the theory that if we cannot solve the easy problem our chances of solving the harder problem are certainly not better.

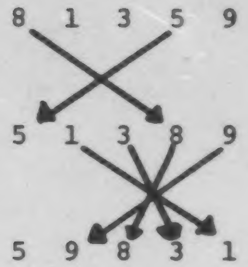
The transformation of 95138 to 93851 by the preceding steps can be represented schematically as



Other transformations appear as







To insure that the procedure does indeed produce the next smaller number in the list, it is instructive for the student to choose other numbers and construct the schematics for their transformation using the above steps.

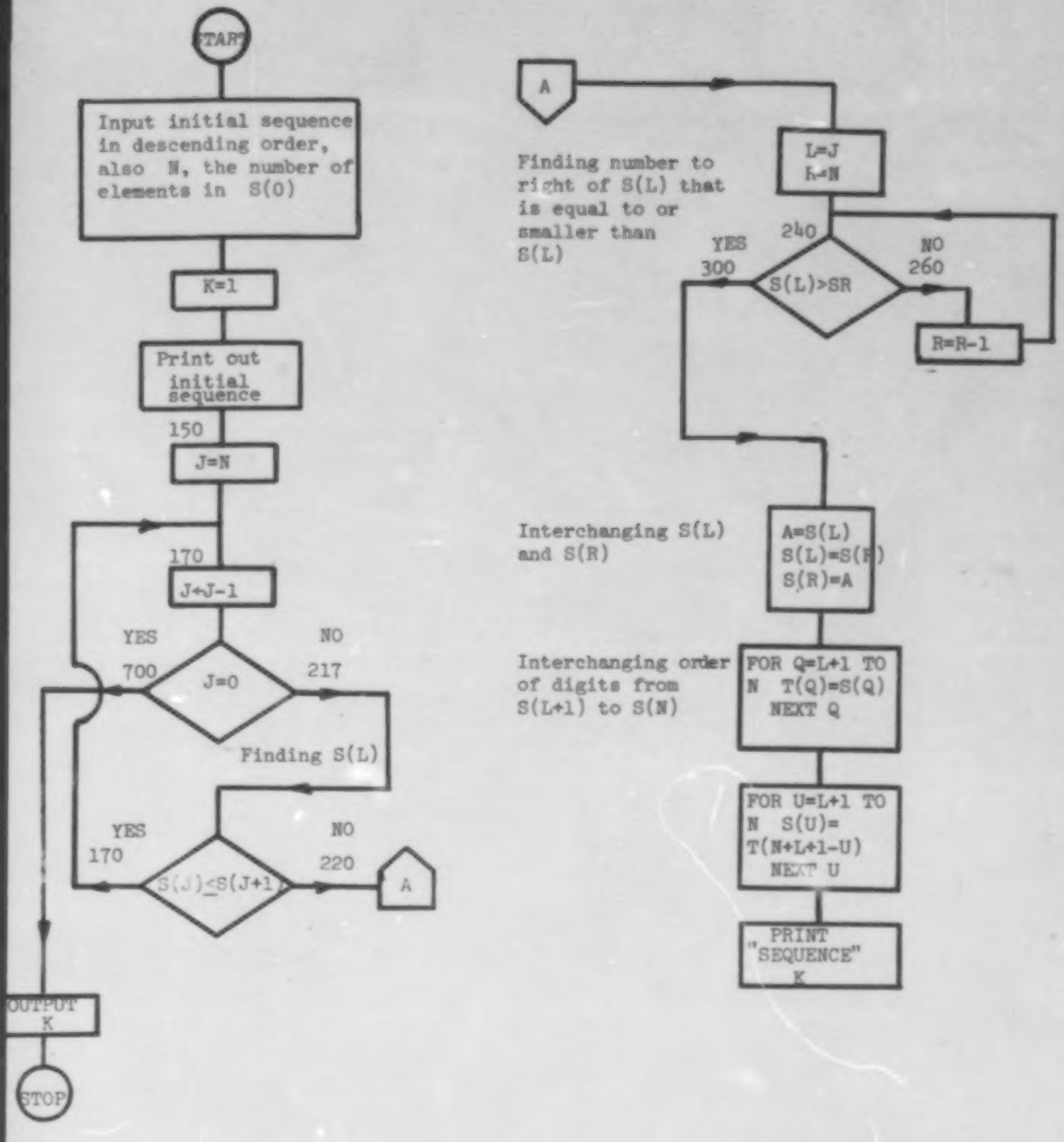
The determination of the digit to the right of S(L) that is to be interchanged with S(L) is based upon the fact that all of the digits to the right of S(L) are in increasing order; that is, S(L+1), S(L+2), S(L+3), ..., S(N) are a non-decreasing sequence of digits. This means that in order to find the next smaller digit than S(L) we have merely to search for the first digit to the right of S(L) that is smaller than S(L). Thus if S(R) denotes the first digit to the right of S(L) such that S(L)>S(R) the digit S(R) is the digit to be interchanged with the digit S(L). To minimize the number of operations the search procedure is made from right to left. See instruction 240 or 260.

The basis for the transformation has now been described and the interchanging is repeated until the digits are all in ascending order. The program as written will accept numbers up to 20 digits. There are many BASIC dialects and it may be the case that the computer accessed by the reader has a dialect which permits easier entry of the initial number. Should this be the case the reader can easily modify the program. A flowchart of the process is shown in figure 8.8 and the program is listed in figure 8.9. Figure 8.10 accompanying the program listing, is a typical output listing the number of distinct combinations of choosing 3 recovered cows from 5 cows. The notation is that used in the discussion.

The printing of all of the combinations can be suppressed by making the following modifications to the program.

1. Eliminate lines 105, 110, 112, 113, 114, and 115
2. Change line 570 to read, 570 GOTO 150

By increasing the number 1000 appearing in lines 565 and 705, the program can accommodate problems having more than 1000 distinct combinations.


START
Input initial sequence in descending order, also N, the number of elements in S(0)
K=1
Print out initial sequence
150
J=N
170
J←J-1
YES
700
J=0
NO
217
Finding S(L)
YES
170
S(J)≤S(J+1)
NO
220
A
OUTPUT K
STOP
A
L=J
R=N
Finding number to right of S(L) that is equal to or smaller than S(L)
240
YES
300
S(L)>SR
NO
260
R=R-1
Interchanging S(L) and S(R)
A=S(L)
S(L)=S(R)
S(R)=A
Interchanging order of digits from S(L+1) to S(N)
FOR Q=L+1 TO N T(Q)=S(Q) NEXT Q
FOR U=L+1 TO N S(U)= T(N+L+1-U) NEXT U
PRINT "SEQUENCE" K

Flowchart for Listing Combinations

Figure 8.8

```
1 REM COMBINATION COUNTING
20 DIM S(20),T(20)
50 PRINT "THE NUMBER OF ELEMENTS IN THE SEQUENCE IS"
55 INPUT N
60 PRINT
70 PRINT "THE SEQUENCE MUST BE IN DESCENDING ORDER"
75 PRINT "TYPE IN THE SEQUENCE"
80 FOR I=1 TO N
82 INPUT S(I)
83 NEXT I
85 PRINT
90 PRINT
100 K=1
105 PRINT "SEQUENCE";K
110 FOR I=1 TO N
112 PRINT S(I);
113 NEXT I
114 PRINT
115 PRINT
150 LET J=N
170 LET J=J-1
200 IF J=0GO TO 700
210 IF S(J)<=S(J+1)GO TO 170
220 LET L=J
225 REM   L POINTS TO THE LEFTMOST NUMBER TO BE SWITCHED
230 LET R=N
235 REM  R POINTS TO THE NEXT NUMBER EQUAL TO OR SMALLER THAN S(L)
240 IF S(L)>S(R)GO TO 300
260 LET R=R-1
270 GO TO 240
295 REM   INSTRS. 300 TO 320 INTERCHANGE S(L) AND S(N)
300 LET A=S(L)
310 LET S(L)=S(R)
320 LET S(R)=A
330 GO TO 450
445 REM  INSTRS. 450 TO 500 SWITCH ORDER OF L+1 TO N DIGITS
450 FOR Q=L+1 TO N
460 LET T(Q)=S(Q)
470 NEXT Q
480 FOR U=L+1 TO N
490 LET S(U)=T(N+L+1-U)
500 NEXT U
560 LET K=K+1
565 IF K=1000GO TO 705
570 GO TO 105
700 PRINT "THERE IS A TOTAL OF";K;"COMBINATIONS"
702 GO TO 710
705 PRINT "MORE THAN 1000 COMBINATIONS, PROGRAM AUTAOMATICALLY STOPS"
710 END

READY
```

Figure 8.9

```
THE NUMBER OF ELEMENTS IN THE SEQUENCE IS
?5

THE SEQUENCE MUST BE IN DESCENDING ORDER
TYPE IN THE SEQUENCE
?2
?2
?2
?1
?1


SEQUENCE 1
 2  2  2  1  1

SEQUENCE 2
 2  2  1  2  1

SEQUENCE 3
 2  2  1  1  2

SEQUENCE 4
 2  1  2  2  1

SEQUENCE 5
 2  1  2  1  2

SEQUENCE 6
 2  1  1  2  2

SEQUENCE 7
 1  2  2  2  1

SEQUENCE 8
 1  2  2  1  2

SEQUENCE 9
 1  2  1  2  2

SEQUENCE 10
 1  1  2  2  2

THERE IS A TOTAL OF 10 COMBINATIONS

READY
```

A Typical Result from the Program listed in Figure 8.9

Figure 8.10

PROBLEMS

CHAPTER VIII

1. In a zoo exhibition there are five white tail deer and three mule deer. Three deer are selected at random from the area. What is the probability that all are white tail deer? Two are white tail deer and one is a mule deer? One is a white tail deer and two are mule deer? All three are mule deer?

2. A monkey sits at a typewriter containing the letters A, B, D, G and O. What is the probability of the monkey typing the word BAD? The word GOOD?

3. A laboratory mouse is placed in a maze consisting of five doorways through which it must pass in order to obtain a reward of food. There are four paths leading from one doorway to another and from the start to the first doorway there are three paths. Assuming that no two paths have the same length and also assuming that it is equally likely that the mouse chooses any path, what is the probability that the mouse chooses the sequence of paths whose total length is a minimum?

4. There are three spotted deer in an enclosure each having three, four and five spots respectively.
   (a) What is the probability that a deer selected at random has four spots?
   (b) Two deer are selected at random. What is the probability that the sum of the number of spots is greater than eight?
   (c) A deer is selected at random and the number of spots it has is recorded. The deer is then replaced. A second deer is selected at random and its number of spots recorded. What is the probability that the sum of the number of spots appearing on each deer is greater than eight?

5. In a large city 40% of the population has contacted influenza. What is the probability that in a random sample of 10 people 6 of them have influenza?

6. In an elk herd consisting of 1000 elk, 200 elk are over 5 years of age and 100 of the elk are underfed. An elk is selected at random.
   (a) What is the probability the elk is over 5 years of age? Is underfed?
   (b) What is the probability the elk is over 5 years of age and underfed?
   (c) If the elk is underfed, what is the probability that he is under 5 years of age?

7. In a fish hatchery of 1000 fish, it is noted that 3000 of the fish have become infected with parasite A, 2000 with parasite B, and 1000 of the fish have been under attack from both parasites simultaneously. A fish is chosen at random.
   (a) What is the probability that the fish is subject to attack by parasite A? By parasite B? By both parasites?
   (b) If the fish is under attack by parasite A what is the probability it is under attack by both parasites? If it is under attack by parasite B what is the probability it is under attack by both parasites?

8. Let two fair dice be thrown.
   (a) What is the probability that both numbers are odd?
   (b) What is the probability that only one of the numbers showing is a 3?

(c) What is the probability that the sum is 8?

(d) If the sum is 7 what is the probability one of the numbers is a 2?

(e) If the sum is 7, what is the probability the numbers differ by more than 1?

(f) What is the probability one number is odd and the other even?

9. There are 4 collies, three bulldogs and 5 terriers in a kennel. Four dogs are selected at random.

(a) What is the probability the first two dogs selected are terriers and the third and fourth dogs collies?

(b) If a terrier is selected, what is the probability a bulldog and collie are also selected?

(c) If a terrier is selected on the first draw, what is the probability a bulldog and a collie are also selected?

(d) What is the probability that three bulldogs are selected?

10. Three fair dice are thrown.

(a) What is the probability that their sum is greater than 10?

(b) How many throws must be made so that the probability of of the sum of the numbers being greater than 10 is greater than 80%?

(c) What is the probability that the numbers on the three dice can be arranged in succession.

11. A good mile runner can run the mile in under 4 minutes in 8 out of 10 races whereas a fair mile runner can run the mile in under 4 minutes 2 out of 10 races. A man runs the mile in 3:59. What is the probability he is a fair mile runner?

12. In a corn farm there are twice as many corn plants with tall stalks as short stalks. 60% of the corn with tall stalks produces long ears of corn and 50% of the short stalk corn produces long ears of corn. A stalk of corn is chosen at random.

   (a) What is the probability it is tall and has short eared corn?

   (b) If the stalk has short ears of corn, what is the probability it is a short stalk?

13. Use the data of problem 12. Two stalks of corn are selected at random.

   (a) What is the probability that each stalk is short and produces short corn?

   (b) If both stalks have long ears of corn, what is the probability one stalk is short and the other stalk long?

14. Forest A is four times as large as forest B and both are infected with spruce bud worm. Forest A is 20% infected and Forest B 30% infected.

   (a) A tree is selected at random and found to be infected. What is the probability it came from Forest A?

   (b) Two trees are selected at random and one is observed to be infected and the other healthy. What is the probability the infected tree came from Forest B and the healthy tree came from Forest A?

   (c) What is the probability that three trees selected at random came from Forest A and are healthy?

15. An epidemic breaks out in a population. 30% of the population is infected the first week, 20% is infected the second week and

10% is infected the third week. After the fourth week, there are no new infections. An individual is selected at random in the population and found to be infected. What is the probability the individual was infected in the second week? What is the probability that of three people selected, one was infected each week?

16. A forest is sprayed with DDT which is 70% effective in eliminating a parasite. After spraying, 5 trees are selected.
    (a) What is the probability that all five trees are still afflicted with the parasite?
    (b) What is the probability that three trees are not infected?
    (c) What is the probability that at least three trees are infected?

17. Ten patients are in an operating ward. Five patients have had apendectomies, two have had tonsils removed and three have had ski accidents. Four patients are selected at random.
    (a) What is the probability all four have had their appendix removed?
    (b) What is the probability at least 2 have had their appendix removed?
    (c) What is the probability each of the four selected has had each of the operations?
    (d) If two have had their appendix removed what is the probability that one of the remaining two individuals had his tonsils removed and the other had a ski accident?

18. Ten horses in a racing stable of 15 horses are known to be ill. Four horses are selected.

(a) What is the probability only one is ill?

(b) What is the probability at most one is ill?

(c) If two are ill, what is the probability all four are ill?

19. A hunter is 20% successful in his quest for elk, 30% successful in hunting for white tail deer and 40% successful in hunting for antelope on any given hunting trip.

(a) If he hunts for all three animals on a single trip, what is the probability he will bag all three on a single trip?

(b) What is the probability he will bag all three on five trips?

(c) What is the probability he will get at least one animal in three trips?

REFERENCES

CHAPTER VIII

Uspensky, J. V. 1937. Introduction to Mathematical Probability. McGraw-Hill. New York, NY.

CHAPTER IX

COMPARTMENTAL ANALYSIS

Introduction

This chapter presents the development of a technique, called compartmental analysis, that is useful in describing the flow of a quantity or quantities in biological and ecological systems. Typical examples of such flows are:

(a) The flow of thyroxine as it passes from the blood to the liver and is absorbed into bile.
(b) The flow of energy in an ecosystem.
(c) The exchange of inorganic phosphate between the blood and tissues.
(d) The movement of DDT in a human food chain.
(e) The transfer of organic matter in a peat bog.

Compartmental Analysis is also useful in:

(a) Analyzing rotifer populations viewed as energy transferring systems.
(b) The analysis of enzyme kinetic problems.

The method of presentation will be to discuss examples and to develop the necessary computer programs for their solution. It will be seen that the solution of each example requires the use of the fundamental law of change. Because these problems are usually formulated in terms of the simultaneous ordinary differential equations, we will indicate the connection between the two procedures to enable the student to read the relevant literature.

A fundamental problem in physiology is the determination of the flow of various fluids from one part of the body to another part of the body. This problem is of interest because a knowledge of the transport of fluids within the body is very helpful in determining the function and method of operation of component parts of the body. In the following sections, a very effective technique for analyzing such flows will be described. The technique is called compartmental analysis.

In the literature, the term 'fluid' is usually replaced by the term 'substance'. This replacement occurs because it is frequently quite difficult to identify the particular fluid or substance which is flowing or diffusing into, and/or out of, the compartment of interest. It is known that substances which exist in one part of the body are sometimes transferred or diffused to adjacent parts of the body. However, the mechanism by which the transfer takes place is not always completely understood. One possible explanation is based on the postulation of the existence of a fluid or substance which flows or diffuses from one compartment to another compartment and which transports the substance of interest. This postulation provides the basis for compartmental analysis in physiology.

The analysis consists in modeling the fluid behavior by assuming the existence of compartments into which, and out of which, the fluid flows. Assumptions are then made which specify the amount of fluid which enters and/or leaves a compartment in a small increment of time. These assumptions enable the writing of equations relating the change in the amount of fluid in the compartment in a small increment of time to the amount of fluid entering and/or leaving the compartment in the same unit of time. Thus, it is evident that the fundamental law of change will form the basis of the formulation of the equations. The student who is familiar with hydraulics will note the very close analogy between the method of analysis and that used in hydraulics.

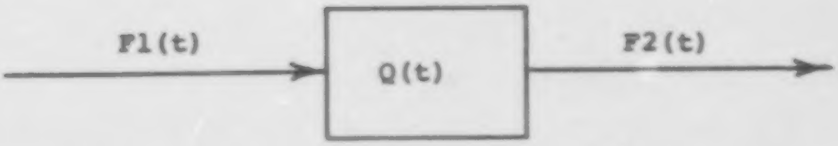
Transfer of the Fluid from the Muscle Tissue

As the first example, we consider the analysis of the transfer of a fluid from the muscle tissue of a guinea pig to the extracellular fluid of the muscle tissue. Since it is very difficult to directly measure the rate of fluid transfer without disturbing the actual flow of the fluid, indirect methods have been developed. Foremost among these methods is the tracer method. This technique consists in labeling the fluid being transferred with radioactive ions. The time variation of the transfer of these ions can then be detected by radioactive sensors. Since it is assumed that the transfer of the ions is caused by their being transported by the flowing or diffusing fluid, the measured rate of transfer of the ions is assumed to be the actual rate of transfer of the fluid. In this way, the effective transfer rate of the fluid can be determined. Before presenting a discussion of the tracer method (see the section entitled, "Tracer Methods"), we illustrate our method of analysis of some simple problems so that the student can more readily appreciate the ideas.

Using the tracer technique, Born and Bulbring (1956), radioactively labeled the fluid with potassium ions and succeeded in analyzing the transfer of these ions from the guinea pig muscle tissue to the extracellular fluid.

In this experiment, the smooth muscle of guinea pigs was permitted to take up $^{42}K^+$ ions until the radioactivity had become constant in the muscle. The muscles were then immersed in a constantly flowing non-radioactive extracellular fluid and the time variation of the concentration of the $^{42}K^+$ in the muscles was recorded. The observed results indicated a decreasing loss in concentration of the potassium ions and hence the ions were being transported from the muscle to the extracellular fluid. It is desired to construct a simple explanation whose quantitative prediction would agree with the empirical results.

This may be done by assuming that the muscle acts like a compartment or a container in which the $^{42}K^+$ ions are deposited. The compartment is filled with potassium ions, and the stored ions are then permitted to diffuse out of the compartment. The process may be conveniently depicted in terms of a flow diagram in which the muscle is imagined to be a compartment or tank into which potassium ions are deposited and from which they are diffusing. The flow of ions into the muscle is denoted by F1(t) and the transfer of ions from the tissue to the extracellular fluid is denoted by F2(t). Both of these flows may, and usually do, vary in time. Q(t) will denote the quantity of the potassium ions in the muscle. Using this notation, the flow diagram of the process appears in figure 9.1 as:

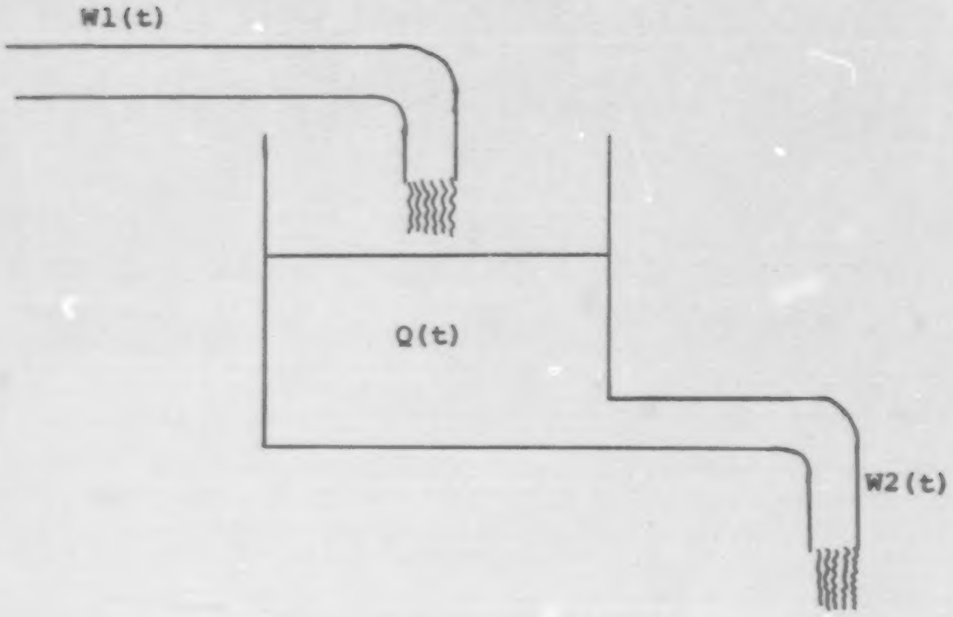


A Single Compartment Model

Fig. 9.1

In this work the flow of a quantity is the rate at which that quantity is adding to, or exiting from, a compartment. Thus, the unit of flow is expressed in volume per unit time or mass per unit time. These units are usually written as volume$\cdot$time^{-1} or mass$\cdot$time^{-1} respectively. The use of the term flow in this work should not be confused with the term mass flow or volume flow as used in hydrodynamics. In this latter usage, these terms designate the transport through space of a specified amount of mass or volume in a unit time. We again remind the student that the use of argument t in the flow variables and the volume variable indicates that, since the quantities vary in time, the quantities are to be evaluated at a specified time, t. In other words, such a notation indicates that the variables are functions of time. In this example, the ions do not return from the extracellular fluid to the tissue. Since we are interested only in the variation in time of the concentration of ions in the tissue, there is no need to depict the extracellular fluid as a compartment. This simplification also means that it is only necessary to depict the ions entering and leaving the "muscle tissue" compartment.

In conceptualizing a model for the transfer of the potassium ions from the tissue to the extracellular fluid, it is helpful to think of a fluid flowing into and out of a tank. A hydrodynamic model analogous to the single compartment system is displayed in figure 9.2:



Hydrodynamic Analogy
Fig. 9.2

where Q(t) denotes the amount of water in the parallel-sided tank and W1(t) and W2(t) denote the respective input and output flow rates. Q(t) is measured in volume units and accordingly the flow rates W1(t) and W2(t) are expressed in units of volume per unit time. Typical volume units for Q(t) are cubic centimeters, liters, gallons, etc., and the corresponding typical units for the flow rates are cubic centimeters per second, liters per minute, gallons per hours, etc. Frequently, the amount Q(t) is expressed in units of mass such as grams or pounds. In this event, the quantities W1(t) and W2(t) are called mass flow rates and are expressed in units of mass per unit time, i.e. grams per second, pounds per minute, etc. In this hydrodynamic analogue the input flow rate is prescribed, as is the cross-sectional area A of the tank, and it is desired to calculate Q(t). This can be accomplished by using the fundamental law of change to write:

$$Q(I+1) = Q(I) + C(I) \tag{9.1}$$

where $C(I)$ denotes the change in the amount of the fluid in the I^{th} time increment. For convenience we will assume that the time increment $D1$ is small and constant and therefore, the actual time, $T(I)$, is given by

$$T(I) = T(0) + I * D1. \tag{9.2}$$

In equation (9.1), $C(I)$ is equal to the change in the amount of fluid in the tank in a time increment. If $E1(I)$ represents the amount of fluid entering the tank and $E2(I)$ denotes the amount of fluid leaving the tank during the time increment, then

$$C(I) = E1(I) - E2(I).$$

Moreover, if $A1$ and $A2$ denote the cross-sectional areas of the respective pipes through which the fluid enters and exits, then

$$E1(I) = F1(I) * D1,$$

and

$$E2(I) = F2(I) * D1.$$

From elementary physics it is known that the rate of the fluid exiting through an orifice of unit cross-sectional area is proportional to the depth (or head) of the fluid in the tank. Thus,

$$F2(I) = K1 * H(I)$$

where H(I) denotes the depth of the fluid in the tank during the I^{th} time increment and K1 is the constant of proportionality. If A3 denotes the cross-sectional area of the tank, we can write

$$Q(I) = A3 * H(I)$$

or

$$F2(I) = K1 * Q(I)/A3.$$

Hence,

$$E2(I) = K1 * Q(I) * D1/A3,$$

and the fundamental law of change may be used to give

$$Q(I+1) = Q(I)+(F1(I)* D1-K1*Q(I)*D1/A3). \tag{9.3}$$

If the ratio K1/A3 is denoted by K, this equation can be simplified to give

$$Q(I+1) = Q(I)+F1(I)*D1-K*Q(I)*D1. \tag{9.4}$$

The term

$$K*Q(I)*D1$$

represents the amount of fluid being transferred from the tank in the time increment, and it is seen that this amount is proportional to the quantity Q(I), of fluid in the tank. The term Fl(I)*Dl represents the amount of fluid entering the tank in the time increment.

We now return to the potassium ion transport problem and note that there is a direct analogue to the relation just derived. This follows from the fundamental assumption concerning the transport of a substance out of a compartment in a small increment of time Dl. The assumption is: "In a small increment of time, the amount of the substance leaving a compartment is proportional to the concentration of the substance in the compartment". This is frequently called Fick's law. Thus, for this particular problem, the law states: "In a small time increment, Dl, the amount of the ionized potassium leaving the muscle tissue compartment to enter the extracellular fluid is proportional to the concentration of the potassium ions in the muscle tissue".

If Q(I) now denotes the amount of ionized potassium at the beginning of the I^{th} time increment, the quantity of ionized potassium leaving the tissue to enter the extracellular fluid during the time Dl may be written as

Kl*Q(I)*Dl.

Kl is the constant of proportionality and is measured in units of reciprocal time. Such constants are called rate constants or transfer coefficients and their determination is of fundamental importance. The amount of ionized potassium entering the tissue in the time increment Dl is Fl(I)*Dl where Fl(I) denotes the rate of flow of the entering ionized potassium at the time t. The equation governing the time variation of the concentration of the ionized potassium in the tissue is obtained by using the fundamental law of change. This gives

$$Q(I+1) = Q(I)+F1(I)*D1-K1*Q(I)*D1. \qquad (9.5)$$

If there is no addition of potassium ions after the start of the experiment,

$$F(I) = 0, \quad I = 1, 2, 3, \ldots..$$

and equation (9.5) may be written as

$$Q(I+1) = Q(I)-K1*Q(I)*D1. \qquad (9.6)$$

This is the equation describing the simple diffusion of the potassium ions from the tissue to the extracellular fluid. Equation (9.5) is just a different form of the equation governing the flow of a fluid from a tank or of the equation describing Malthus population growth. This may be seen by noting that the change in the variable in each of the fundamental equations has the same form. The respective changes in the variables are F1(I)*D1-K1*Q(I)*D1, F1(I)*D1-K*Q(I)*D1 and B*P(I)-M*P(I). Each of these terms is a difference between "what comes in and what goes out" in the time period, and therefore, each is a statement about the conservation of the respective variables. This suggests that there is a general conservation principle governing many diverse phenomena and this is indeed the case. It is possible to use such a principle as a basis for the derivation of the preceding equations. In fact, the equations governing many diverse physical phenomena are usually derived from conservation principles. In this way, conservation principles form a basis for demonstrating the unity of the physical sciences. The occurrence of the same equation under different guises is thoroughly exploited by mathematicians to also illustrate the unity of the disciplines in which the equations occur. In addition, mathematicians use these similarities to extend and derive solutions of the equations.

The student who has studied the calculus will note that a mathematical formulation of this problem is given by

$$\frac{dQ}{dt} = -Kl \cdot Q \qquad (9.7)$$

$$Q(t=0)=Q_0,$$

and the solution is given by

$$Q(t) = Q_0 \exp(-Kl \cdot t). \qquad (9.8)$$

The appearance of the exponential function accounts for the term "exponential decay" with a decay or loss rate of Kl. Since equation (9.7) describes many diverse phenomena, its solution (9.8) is very useful and has been well studied. It is frequently the case that the results obtained from compartmental analysis are better graphically portrayed with the aid of a logarithmic scale. This fact suggests the use of the technique developed in the first chapter for compressing the variation of the concentration to assist the graphical representation of the program results.

There is no acceptable format for the depiction of compartmental models. A frequently encountered format corresponding to the single compartment potassium ion model is shown in figure 9.3 below:

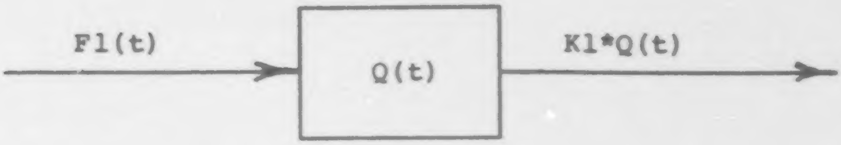


Fig. 9.3

Another format is shown in figure 9.4 below:

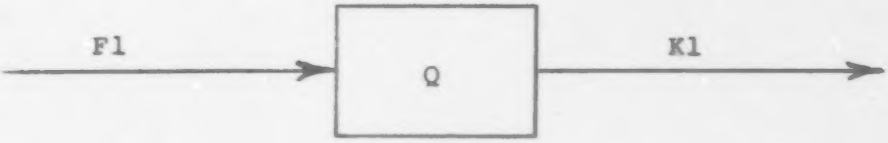


Fig. 9.4

where it is understood that the amount of the substance leaving the compartment in a small increment of time, D1, is equal to the product of the constant of proportionality K1, the concentration Q(t) present in the compartment during the time increment, and D1.

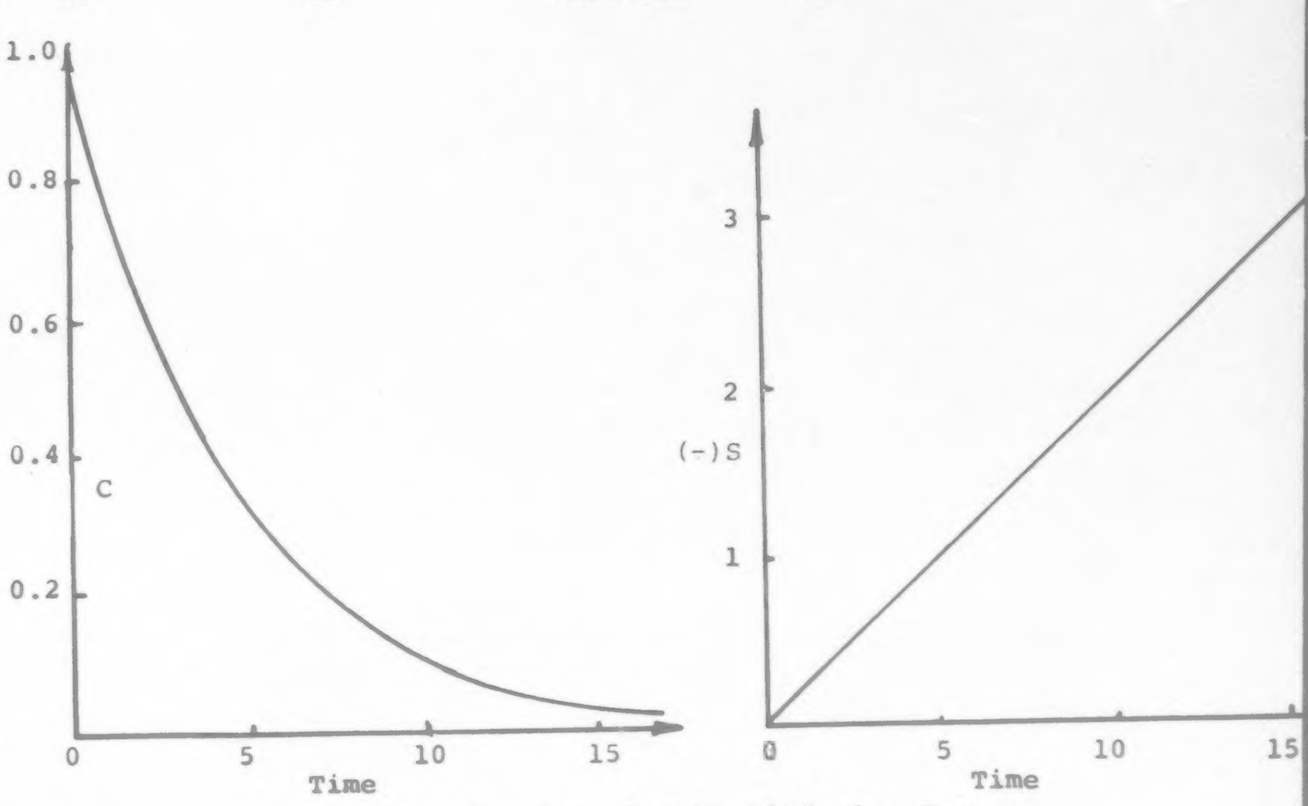
As a numerical example, we consider the loss of $^{42}K^{+}$ ions from the muscle to the extracellular fluid when the muscle has an initial ion concentration of unit strength. The program is shown in figure 9.5 and line 95 of the program specifies the initial strength of the ion concentration in the muscle. The constant of proportionality or rate constant was chosen to be 0.19 min.$^{-1}$, the time increment was set at 1 min. and the program run for 20 time increments for an actual total time of 20 minutes. In the program, C(I) denotes the concentration of the ionized potassium and line 120 expresses the governing equation, equation 9.6. Lines 140 and 145 are inserted merely to illustrate the use of the compressed scale, S(I), and line 100 initializes the value of S(0) to coincide with the actual value of the natural logarithm of the initial concentration. The elapsed time is calculated in lines 105 and 130. The results are portrayed in both tabular and graphical form in figure 9.6.

```
1 REM    THE SIMPLE DIFFUSION OF POTASSIUM IONS,  EQ. 9.6
10 DIM C(105),T(105),S(105)
20 REM  C(I)  DENOTES THE ION CONCENTRATION
22 REM  T(I)  DENOTES THE ELAPSED TIME TO THE  ITH TIME INCREMENT
24 REM  S(I)  DENOTES THE ION CONCENTRATION ON A COMPRESSED SCALE
50 PRINT "TYPE THE TRANSFER COEFFICIENT  K1  AND THE TIME INCREMENT  D1"
60 INPUT K1,D1
65 PRINT
70 PRINT "TYPE THE NUMBER OF TIME INTERVALS, N"
80 INPUT N
85 PRINT
86 PRINT
94 REM  C(0)=1  CORRESPONDS TO UNIT STRENGTH INITIAL ION CONCENTRATION
95 LET C(0)=1
99 REM  INITIALIZING  S(0)  TO CORRESPOND TO  LOG(C(0))
100 LET S(0)=LOG(C(0))
104 REM  EXPERIMENT STARTS AT TIME  T=0
105 LET T(0)=0
110 FOR I=0 TO N
119 REM  INSTR.  120  IS THE FUNDAMENTAL SIMPLE DIFFUSION EQUATION
120 LET C(I+1)=C(I)-K1*D1*C(I)
125 REM   INSTR. 130 CALCULATES THE ELAPSED TIME
130 LET T(I+1)=T(I)+D1
135 REM LINES 140 &145 CALC. THE COMPRESSED SCALE VALUES OF THE CONC.
140 LET R=(C(I+1)-C(I))/C(I)
145 LET S(I+1)=S(I)+R
150 NEXT I
180 PRINT "I             T(I)            C(I)         S(I)"
185 PRINT
190 FOR I=0 TO N
210 PRINT I,T(I),C(I),S(I)
220 NEXT I
230 END

READY
```

Simple Diffusion Program

Fig. 9.5

```
TYPE THE TRANSFER COEFFICIENT  K1  AND THE TIME INCREMENT  D1
?0.19, 1

TYPE THE NUMBER OF TIME INTERVALS, N
?20

I                  T(I)              C(I)           S(I)

 0                  0               1              0
 1                  1               .81           -.19
 2                  2               .6561         -.38
 3                  3               .531441       -.57
 4                  4               .430467       -.76
 5                  5               .348678       -.95
 6                  6               .28243        -1.14
 7                  7               .228768       -1.33
 8                  8               .185302       -1.52
 9                  9               .150095       -1.71
 10                 10              .121577       -1.9
 11                 11              .0984771      -2.09
 12                 12              .0797664      -2.28
 13                 13              .0646108      -2.47
 14                 14              .0523348      -2.66
 15                 15              .0423912      -2.85
 16                 16              .0343368      -3.04
 17                 17              .0278128      -3.23
 18                 18              .0225284      -3.42
 19                 19              .018248       -3.61
 20                 20              .0147809      -3.8
```



Results From Simple Diffusion Program

Figure 9.6

In developing the expressions for the amount of the substance entering or exiting from a compartment in a small increment of time, we postulated that this amount could be represented as a flow rate multiplied by the small increment of time. However, it was also stated that the flow rate varied in time. Now the critical reader might well ask about the accuracy of statements like, "the amount of the substance entering the compartment in the small time increment, Dl, is equal to Fl(I)*Dl where Fl(I) is the entering flow rate at the beginning of the time increment". Such a statement is not precise, that is, it is not rigorously true; however, if Dl is sufficiently small, the statement is "very nearly true". An alternative argument for justifying our approach is noted that if the flow rate changes smoothly in time, by making Dl sufficiently small, the error committed in using the expression Fl(I)*Dl to represent the amount of fluid entering the compartment in the time increment, can be made as small as is desired. The student who is familiar with the calculus, will recall that an application of the mean value theorem can be used to justify the assertion.

In essence we are saying there is very little or negligible error incurred by assuming that Fl(t) does not change during the time increment Dl. For this reason, it is convenient to speak of the value of Fl(t) during the time increment as the value of Fl(t) at the beginning of the time increment. Similar comments are to be made about other variables which change during the entire time of interest. It is important to note that by making the above assumptions, we are not restricting the variation of the flow rate from one time increment to another time increment, rather we are assuming that the flow rate is constant during the time increment. Graphically, we are replacing a smooth or continuous curve by a set of steps and using a "step like" approximation to the curve.

Since the amount of fluid is continuously changing, the rigorous analysis of such flows would require the use of differential equations. If the problem is described by differential equations but becomes at all complicated, a digital computer must be employed to effect a solution. The resultant computer based solution is equivalent to dividing the time of interest into small time increments and assuming that the flows are constant in each time increment. This latter procedure is analogous to our procedure. However, in fairness it should be pointed out that very accurate methods have been developed for the numerical solution of differential equations. The formulation that we have used, and will continue to use in this chapter, is equivalent to a first order Euler method of solution. We will continue to use this method of presentation because it permits an easier concentration on the formulation of the problems and the development of programs. This is the primary purpose of the course. More accurate methods of solution of the model equations will be left to other more advanced courses.

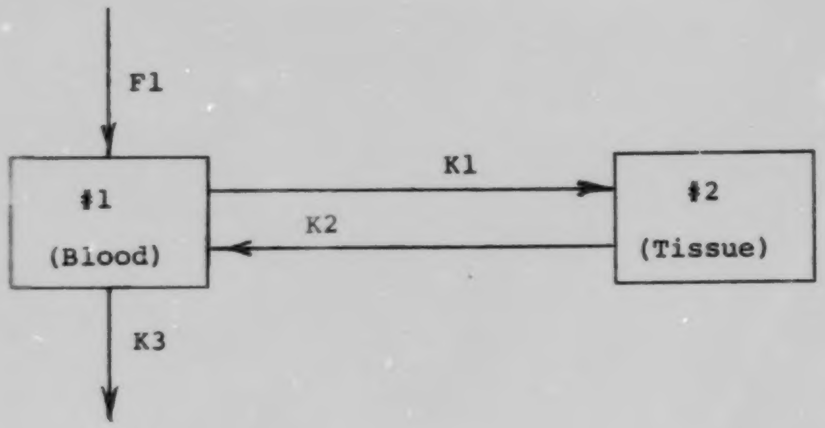
In case the student feels he is being given an incorrect formulation, it should be pointed out that the fundamental model is itself a very crude approximation to the actual biological phenomena even if it is formulated in the language of mathematics. Elegant and sophisticated analysis of an incomplete or irrelevant model is far less satisfactory than an approximate analysis of a realistic model. In addition, the experimental and data collection errors associated with the obtaining of the necessary data frequently mitigate against an overemphasis on the numerical accuracy of the solution of a prescribed set of model equations.

Example #2

This example develops a simple compartmental model for the kinetics of the transfer of inorganic phosphate in a physiological system over a period of time. It is assumed that inorganic phosphate enters the blood from some source and that a certain portion enters the tissue surrounding the blood and the remainder goes elsewhere. It is further assumed that the phosphate entering the tissue is returned to the blood at a different rate than it entered the tissue. The problem is to describe the time variation of the internal exchange of the inorganic phosphate between both the blood and the tissue.

The discussion of this example will be more complete and will include a delineation of the assumptions, and some of the restrictions, of compartmental analysis.

Let the blood be designated as compartment 1, and the tissue as compartment 2. A diagramatic representation of the two compartment system is shown in figure 9.7.



Two Compartment Model of Transfer
of Inorganic Phosphates

Fig. 9.7

The diagram is an oversimplification of the actual process by which the inorganic phosphate enters the blood, is transferred to the tissue and is then returned to the blood. Many complex biochemic processes are involved in the transfer. For example, the inorganic phosphate could enter the blood from several sources and with differe flow rates from each source. Since an accurate accounting of the different processes by which each source contributes inorganic phosphate to the blood is far too complicated, the mass flow of the phosphate into the blood is assumed to be some average of the total of the flows of inorganic phosphate entering the blood. This flow is denoted by F1. A similar remark obtains about the flow of the phosphate leaving the blood. There are several avenues by which the phosphate may leave the blood and the exiting mass flows of each avenue are different. It is assumed that the total of these mass flows can be represented by an average exiting flow F3. It is further assumed that the exiting average mass flow is proportional to the concentration of the inorganic phosphate in the blood.

The mass flow at which the phosphate leaves the blood to enter the tissue is also assumed to be a total mass flow since there is more than one process by which such a transfer of phosphate may occu An analogous statement may be made about the flow of the inorganic phosphate from the tissue to the blood. The constants of proportionality associated with each of the flows are different because the biochemical processes for each flow are different. The technique of representing the total of all contributing flows from one compartment to another by a single total such flow is fundamental to compartmenta analysis since it eliminates the necessity of accounting for each of the contributing subsidiary flows. Because of this, compartmental analysis is a great oversimplification of the total chemical kinetic process. A more thorough accounting of each individual flow, togethe with a determination of the fundamental chemical kinetics necessary to understand such mass transfers, is the subject of much intensive investigation. We have used the term mass flow because we are interested in describing the mass of the inorganic phosphate in various components of the system at different instants of time. This require the notion of a flow of mass or mass flow rate which, as was previousl noted, is measured in units of mass · time^{-1}. A complete descriptio of the dynamics of the exchange of the inorganic phosphate between the blood and the tissue requires a knowledge of the time variation of the quantity of inorganic phosphate in both the blood and the tissue. The direct experimental obtainment of this variation is usually not possible and consequently indirect methods must be used. As mentioned previously, one such method is the tracer method which will be discussed later. Because this example involves the determination of the time variation of quantities of inorganic phosphate in the two compartments, the fundamental quantities of interest are the masses of phosphate in the various compartments. In many problems, such as those involving the determination of the intensity of radiation or the chemical kinetics of a reaction, the concentration of a substance or of a material is the primary quantity of interest. Compartmental analysis is also a useful method for the analysis of these problems.

With this heuristic discussion as a background, we list some of the simplifying assumptions that are usually made in applying compartmental analysis.

(1) The amount of substance being transferred from a compartment during a short period of time is proportional to the concentration of the substance in the compartment and to the length of the time increment Dl. Since the volume of the compartment is assumed to be constant, this assumption implies that the amount of the substance leaving a compartment in a time increment is proportional to the mass of the substance in the compartment. There is abundant experimental evidence substantiating the validity of this assumption, e.g. Sheppard (1962). The constant of proportionality is frequently called a transfer coefficient, because it relates the effective amount of the substance actually being transferred in the small time increment to the amount of the substance present in the compartment at the time of the transfer.

(2) The molecules of the reaction participate in the process in a random way. Thus, there is no distinction between the old and the newly formed molecules.

(3) The kinetic processes are irreversible.

(4) The system being analyzed is described by the fewest number of possible compartments.

(5) The compartments are of constant size throughout the time of interest. Thus, if the concentration of the substance is known, the amount of the substance in the compartment is

obtained by multiplying the concentration by the volume of the compartment. In this way the time variation of the quantity of the substance in the compartment may be obtained if the time variation of the concentration is known. Analogously, if the time variation of the total quantity in the compartment is known, the variation of the concentration may be obtained by dividing by the volume of the compartment.

(6) The material, upon entering the compartment, is instantaneously thoroughly stirred and mixed. In reality, there is a short but finite time of mixing; but this approximates instantaneous mixing if the mixing time is very short compared to the turnover time of the mass of material in the compartment.

These assumptions are modifications of those given in Atkins (1969) and the serious student is encouraged to become familiar with this work. The first and last assumption form the basis of the compartmental analysis of the tracer method. Their importance will be more fully appreciated after the student has read the section on the determination of constants. It is customary to state the first assumption in terms of the instantaneous time rate of change. Thus, the first assumption is usually stated as, "The rate of loss of a substance from a compartment is proportional to the concentration of the substance in that compartment". Since, for short increments of time, the loss rate may be assumed to be constant, our statement of the first assumption in terms of the transfer of an amount of material in an increment of time is a simple modification of the usual statement. Assumption #1 is very important as it permits the calculation of the amount of substance being transferred in an increment of time. Recent theoretical efforts in compartmental analysis are aimed at permitting the removal of some or all of these assumptions. The discussion of such methods requires a very detailed discussion of the biochemistry and physiology or biology peculiar to the phenomenon under analysis and consequently will not be discussed here. This fact is mentioned to indicate another area in which computer assisted analysis is enlarging understanding.

In the compartment diagram shown in fig. 9.7, F1 designates the mass flow of the inorganic phosphate from an outside source into the blood. This flow is prescribed and is analogous to a driving force in a mechanical or physical system. It may vary in time or be constant. The constant of proportionality relating the amount of inorganic phosphate which leaves the blood in an increment of time to the amount of phosphate in the blood during this time increment is denoted by K1. Similarly, K2 denotes the transfer coefficient which relates the amount of phosphate leaving the tissue to enter the blood during the time D1 to the amount of inorganic phosphate in the tissue during that time increment. F3 designates the increment of mass of the phosphate leaving the blood to go elsewhere and the corresponding constant of proportionality is denoted by K3. Thus,

$$F3(I) = K3*Q2(I)*D1.$$

If C1(I) and C2(I) denote the concentration of the phosphate in the blood and the tissue respectively, and Q1(I) and Q2(I) denote the corresponding masses of phosphate, we have

$$Q1(I) = V1*C1(I)$$

and

$$Q2(I) = V2*C2(I)$$

where V1 and V2 denote the respective volumes of the compartments.

The equations of compartmental analysis are frequently derived using the Fick principle which is a statement of the conservation of material in the system. The principle is a statement governing the time behavior of a quantity or substance, S, being transported by the fluid. The principle states that: "In a small increment of time, the change in the quantity S in a compartment is equal to the difference between the amount of S entering the compartment and the amount of S leaving the compartment in the same increment of time plus the amount of S being created in the compartment minus the amount of S being destroyed in the compartment during this same time increment". In all of the examples with which we shall be concerned, it will be assumed that there is no creation nor destruction of the substance S in the compartment and consequently we shall limit our discussion to those problems requiring only an accounting of the entering and exiting flows. With this background we are now prepared to derive the equations describing the flow of the material in the system.

By using the first assumption it is evident that the amount of phosphate transferred from the blood to the tissue in the I^{th} time increment is K1*Q1(I)*D1 and that the amount of phosphate flowing into the first compartment from the tissue is K2*Q2(I)*D1. During the same time increment the amount of phosphate entering compartment #1 from the outside is F1*D1 and the amount leaving the compartment is K3*Q1(I)*D1. In these expressions, D1 represents the length of the I^{th} time increment. In this work, it is assumed that the time increment is constant and small. Now, an application of Fick's principle to the first compartment gives

$$Q1(I+1) = Q1(I)+(F1(I)-K1*Q1(I)+K2*Q2(I)-K3*Q1(I))*D1. \tag{9.9}$$

When applied to the second compartment, the Fick principle yields

$$Q2(I+1) = Q2(I)+(K1*Q1(I)-K2*Q2(I))*D1. \tag{9.10}$$

We remind the student that the quantities $Q1(I+1)-Q1(I)$ and $Q2(I+1)-Q2(I)$ are the respective differences of the concentrations in each compartment from one period to the next. Thus, the previous discussion was concerned with the development of expressions for the change in the variables in one time period in order to apply the fundamental law of change. By now the student should have realized that the discussion of most of the topics in this work has been concerned with the obtainment of the change in the variable or variables in a time period in order to use the law of change. This reminder is being given in order that the student might better appreciate the simplicity of the approach and the utility of the law.

The construction of the principle part of the program is quite simple. Provision must be made for the entering of the transfer coefficients. The entering of the time variation of inflowing inorganic phosphate, $F1(I)$, may be accomplished by a table or a function. The program was run assuming an initial concentration of inorganic phosphate in the blood equal to unit strength. It was further assumed that no phosphate entered the blood except that being returned from the tissue. Thus, $F1(I)$ was set equal to zero for the entire run. The values for the constants of proportionality were chosen identical to those used by Atkins (1969) and runs were made corresponding to a time increment of 1 unit. Figure 9.8 lists the program and it should be self-explanatory. The heart of the program is contained in lines 200 to 230. Figure 9.9 displays the output of a typical run, and figure 9.10 portrays the graphical results of a typical run.

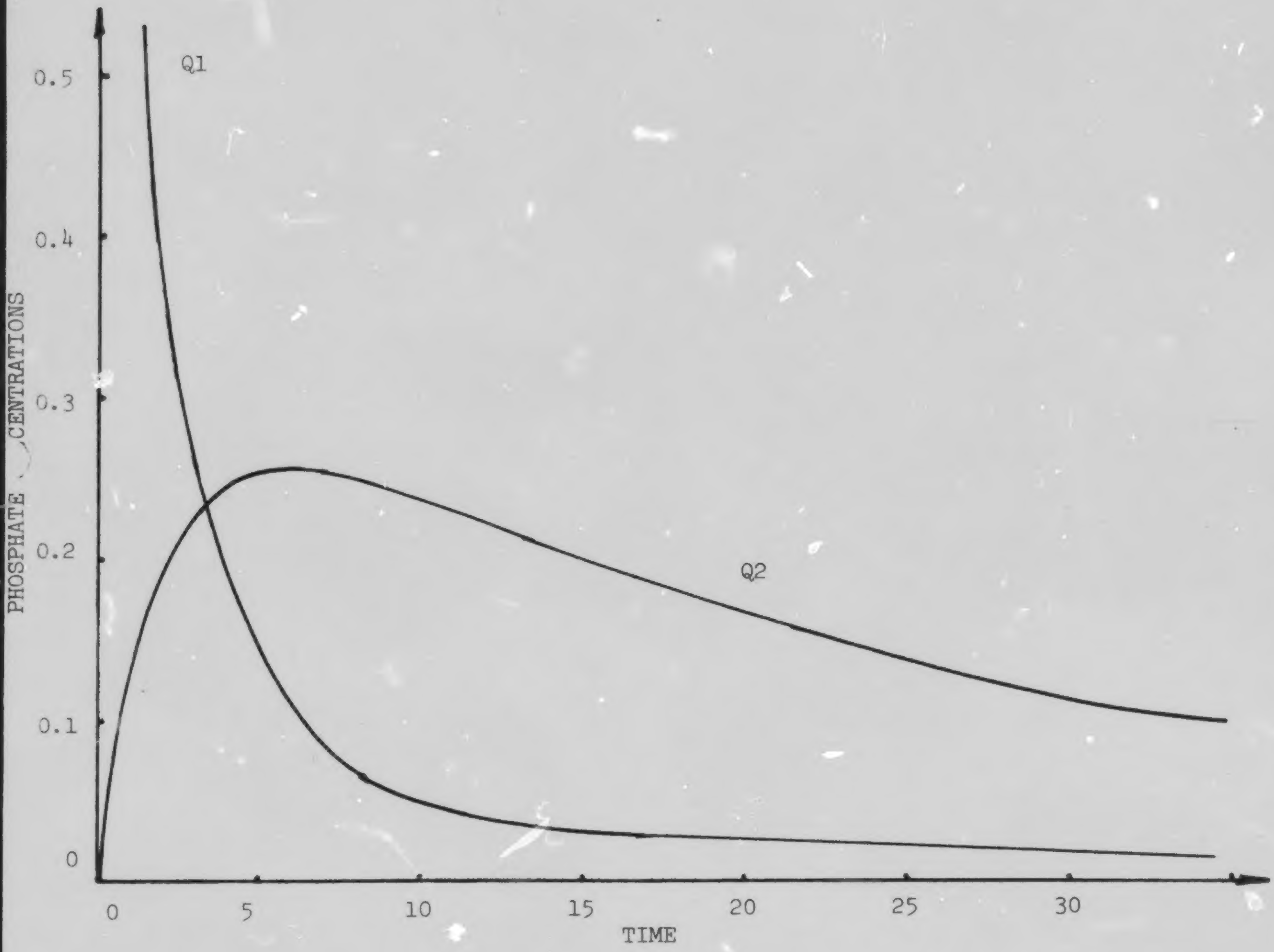
```
1 REM            JULY 30, 1975
5 REM   TWO COMPARTMENT MODEL OF TRANSFER OF INORGANIC PHOSPHATE
10 DIM Q1(205),Q2(205),F1(205),T(205)
20 PRINT "TYPE THE TRANSFER COEFFICIENTS K1, K2, K3"
30 INPUT K1,K2,K3
35 PRINT
40 PRINT "TYPE THE TIME INCREMENT  D1"
45 INPUT D1
46 PRINT
47 PRINT
48 PRINT
55 FOR I=0 TO 200
57 REM  INST. 55 TO 70 ASSURE NO INFLOWING PHOSPHATE
60 LET F1(I)=0
70 NEXT I
185 T(0)=0
187 REM  SETTING Q1(0)=1  GIVES UNIT INITIAL STRENGTH
190 LET Q1(0)=1
195 LET Q2(0)=0
197 REM  INSTR. 200 TO 230 ARE PRINCIPLE PART OF PROGRAM
200 FOR I=0 TO 200
205 LET T(I+1)=T(I)+D1
210 LET Q1(I+1)=Q1(I)+(F1(I)-K1*Q1(I)+K2*Q2(I)-K3*Q1(I))*D1
220 LET Q2(I+1)=Q2(I)+(K1*Q1(I)-K2*Q2(I))*D1
230 NEXT I
290 PRINT " I           T(I)          Q(I)         Q2(I)      F1(I)"
295 PRINT
300 FOR I=0 TO 40
320 PRINT I,T(I),Q1(I),Q2(I),F1(I)
330 NEXT I

READY
```

Fig. 9.8

```
TYPE THE TRANSFER COEFFICIENTS K1, K2, K3
?.118, .056, .234

TYPE THE TIME INCREMENT  D1
?1


 I             T(I)           Q(I)          Q2(I)        F1(I)

 0              0              1             0              0
 1              1             .648          .118            0
 2              2             .426512       .187856         0
 3              3             .2869         .227664         0
 4              4             .19866        .248769         0
 5              5             .142663       .25828          0
 6              6             .106909       .260651         0
 7              7             .0838736      .25867          0
 8              8             .0688356      .254081         0
 9              9             .058834       .247975         0
 10             10            .0520111      .241031         0
 11             11            .0472009      .233671         0
 12             12            .0436718      .226155         0
 13             13            .040964       .218643         0
 14             14            .0387887      .211233         0
 15             15            .0369641      .203981         0
 16             16            .0353757      .19692          0
 17             17            .033951       .190067         0
 18             18            .032644       .183429         0
 19             19            .0314253      .177009         0
 20             20            .0302761      .170805         0
 21             21            .029184       .164812         0
 22             22            .0281407      .159027         0
 23             23            .0271407      .153442         0
 24             24            .0261799      .148052         0
 25             25            .0252555      .14285          0
 26             26            .0243651      .13783          0
 27             27            .0235071      .132987         0
 28             28            .0226799      .128314         0
 29             29            .0218821      .123804         0
 30             30            .0211127      .119453         0
 31             31            .0203704      .115255         0
 32             32            .0196543      .111205         0
 33             33            .0189635      .107296         0
 34             34            .0182969      .103525         0
 35             35            .0176538      .0998871        0
 36             36            .0170334      .0963766        0
 37             37            .0164347      .0929894        0
 38             38            .0158571      .0897213        0
 39             39            .0152998      .086568         0
 40             40            .0147621      .0835256        0

READY
```

Output from Program Shown in Fig. 9.8, D1=1.0

Fig. 9.9



Graphical Results from the Program Shown in Fig. 9.8

Fig. 9.10

In order to show the effects on the results of different time increments, the program was modified to accommodate smaller time increments by properly altering the dimension statement and lines 55, 200 and 300. Figures 9.11 and 9.12 show the results for time steps of 0.5 and 0.1 respectively. A comparison of the results corresponding to the three different time increments reveals small differences in the numerical values. For example, the maximum value of the concentration of the inorganic phosphate in the tissue occurs at slightly different times for the three time increments. Similarly, the numerica values of the maximum values are also different as are the values of the phosphate concentrations in both the blood and the tissue for corresponding times. These, and other differences in the results, are due to the magnitude of the time steps. In some problems, it is the case that for a small change in the time step there is an unusually large change in the numerical results. Such problems are said to be unstable. It is important for the student to know of the existence of instability and to be able to recognize it. A crude method for detecting the existence of instability is to compare the degree of agreement of results obtained by running the program with time increments which are within 10 to 25 per cent of each other. If the results are not reasonably close to one another, there is cause to suspect an instability. These statements are very loose and cavalier and certainly need to be made more precise. The principle concern in this work is to develop computer based methods of analysis. However, it serves no purpose to develop algorithms or programs which are unstable and hence useless. Thus, your author's purpose in mentioning the phenomena of instability at this point is to alert the student to its existence and to suggest a very rough and ready, albeit not infallible, method for detecting it. The effect of accumulating a small error is shown in the second column of figure 9.12 where a non-integer representation of the total time occasionally appears. This error is caused by the computer repeatedly adding the rounded binary representation of 0.1.

With this brief digression we will continue the discussion of compartmental analysis.

```
TYPE THE TRANSFER COEFFICIENTS K1, K2, K3
?.118, .056, .234

TYPE THE TIME INCREMENT  D1
?.5


I        T(I)       Q(I)      Q2(I)    F1(I)

0          0           1           0          0
2          1          .680628     .105964     0
4          2          .468583     .17241      0
6          3          .327601     .212828     0
8          4          .233677     .236142     0
10         5          .170922     .248255     0
12         6          .128819     .253069     0
14         7          .100404     .253164     0
16         8          .0810688    .250242     0
18         9          .0677619    .245429     0
20         10         .0584628    .239463     0
22         11         .0518335    .232831     0
24         12         .046988     .225852     0
26         13         .043339     .218734     0
28         14         .0404974    .21161      0
30         15         .0382051    .204567     0
32         16         .0362907    .197658     0
34         17         .0346403    .190916     0
36         18         .0331779    .18436      0
38         19         .0318529    .178001     0
40         20         .0306313    .171842     0
42         21         .0294901    .165883     0
44         22         .0284137    .160122     0
46         23         .0273914    .154556     0
48         24         .0264157    .14918      0
50         25         .0254812    .143989     0
52         26         .0245842    .138976     0
54         27         .0237215    .134137     0
56         28         .022891     .129466     0
58         29         .0220908    .124957     0
60         30         .0213195    .120604     0
62         31         .0205756    .116403     0
64         32         .019858     .112349     0
66         33         .0191657    .108435     0
68         34         .0184977    .104658     0
70         35         .0178531    .101012     0
72         36         .017231     .0974931    0
74         37         .0166306    .0940968    0
76         38         .0160512    .0908189    0
78         39         .015492     .0876551    0
80         40         .0149523    .0846015    0
```

Output from Program Shown in Fig. 9.8, D1=0.5

Fig. 9.11

```
TYPE THE TRANSFER COEFFICIENTS K1, K2, K3
?.118, .056, .234

TYPE THE TIME INCREMENT  D1
?.1


 I             T(I)            Q(I)           Q2(I)        F1(I)

 0             0               1              0                0
 10            1              .701258        .0983711          0
 20            2              .496356        .162241           0
 30            3              .355648        .202635           0
 40            4              .258861        .227088           0
 50            5              .19213         .240748           0
 60            6              .145972        .247134           0
 70            7              .113901        .248647           0
 80            7.99999        .0914822       .246927           0
 90            9              .0756804       .243091           0
 100           10             .0644201       .2379             0
 110           11             .0562814       .231871           0
 120           12             .0502926       .225355           0
 130           13             .0457887       .218588           0
 140           14             .0423144       .21173            0
 150           15             .0395578       .204887           0
 160           16             .0373053       .198128           0
 170           17             .0354102       .191499           0
 180           18             .0337717       .185028           0
 190           19             .0323207       .178733           0
 200           20             .0310092       .172621           0
 210           21             .0298042       .166699           0
 220           22             .0286827       .160966           0
 230           23.0001        .0276286       .15542            0
 240           24.0001        .0266305       .150059           0
 250           25.0001        .0256804       .144879           0
 260           26.0001        .0247722       .139874           0
 270           27.0001        .0239017       .13504            0
 280           28.0001        .0230656       .130372           0
 290           29.0001        .0222613       .125864           0
 300           30.0001        .0214869       .121512           0
 310           31.0001        .0207406       .117309           0
 320           32.0001        .020021        .113252           0
 330           33.0001        .019327        .109334           0
 340           34.0001        .0186575       .105552           0
 350           35             .0180114       .101901           0
 360           36             .0173879       .0983762          0
 370           37             .016786        .094973           0
 380           38             .0162051       .0916876          0
 390           39             .0156444       .0885158          0
 400           40             .0151031       .0854536          0
```

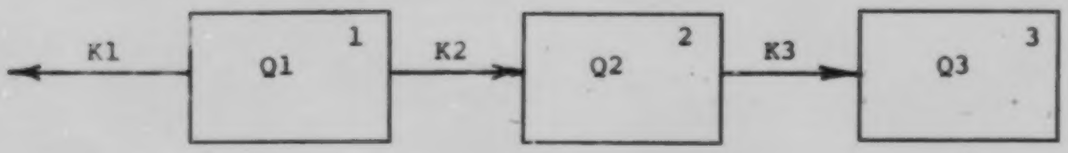
Output from Program Shown in Fig. 9.8, D1=0.1

Fig. 9.12

Example #3

This example is the study of the metabolism and excretion of paracetamol by Cummings, King and Martin (1967) and is taken from Atkins (1969). Their experiment consisted in measuring the elimination of some of the metabolites of paracetamol in a human who had presumably received the drug orally. Following Atkins we will consider only the formation and excretion of paracetamol sulphate.

It is thought that the paracetamol enters the plasma and that it is in the plasma where the metabolites of paracetamol are formed. A portion of the paracetamol is transformed into paracetamol sulphate and is then transferred to the urine where the time variation of the sulphate was experimentally measured. Part of the remaining paracetamol in the plasma is transformed into other metabolites. These metabolites, along with the paracetamol itself, diffuse out of the plasma at a rate which is assumed to be proportional to the quantity of paracetamol present in the plasma. A representation of the compartmental model of this system is shown in fig. 9.13:



Portrayal of the Metabolism and Excretion of Paracetamol
Fig. 9.13

where Q1 denotes the quantity of paracetamol in the plasma and Q2 and Q3 denote the quantity of paracetamol sulphate in the plasma and the urine respectively. In this model the first compartment represents that portion of the plasma which contains the paracetamol which is the source of the sulphates. The second compartment represents that portion of the plasma which contains the sulphates and the third compartment denotes the urine into which the paracetamol sulphates flow. In accord with the assumptions listed above, the paracetamol sulphate is assumed to be generated in the first compartment and to then leave this compartment at a rate which is directly proportional to the quantity of paracetamol in the compartment. K1 will denote the constant of proportionality relating the amount of paracetamol and paracetamol metabolites leaving the plasma in a time D1, and K2 will denote the constant of proportionality relating the amount of paracetamol converted to paracetamol sulphate in the time D1. Finally, K3 will denote the transfer coefficient associated with the transferring of the paracetamol sulphate from the plasma to the urine.

A direct comparison of this development with the mathematical development given by Atkins requires a careful recognition of the different notations. The necessity for the different notations is due to the restriction of a single capital letter followed by a single integer as required in the BASIC programming language.

The derivation of the governing equations is accomplished by applying Fick's principle to each compartment and using the first assumption to enable the application of the law of change to each quantity. For the first compartment it is evident that the quantity

$$(K1*Q1(I)+K2*Q1(I))*D1$$

represents the amount of paracetamol leaving the first compartment in the I^{th} time increment. Thus, the equation governing the change in the amount of paracetamol is

$$Q1(I+1)=Q1(I)-(K1*Q1(I)+K2*Q1(I))*D1. \tag{9.11}$$

The amount of paracetamol sulphate being created in the interval of time D1 is

$$K2*Q1(I)*D1$$

and the amount leaving in the same time interval is

$$K3*Q2(I)*D1.$$

Hence, for the second compartment,

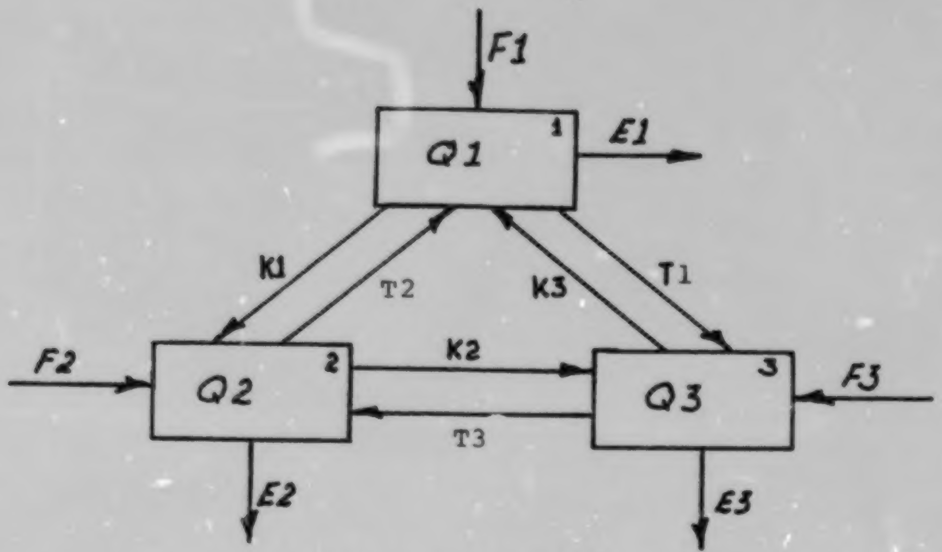
$$Q2(I+1)=Q2(I)+(K2*Q1(I)-K3*Q2(I))*D1. \tag{9.12}$$

In a similar manner

$$Q3(I+1)=Q3(I)+K3*Q2(I)*D1 \tag{9.13}$$

is the equation governing the time change of the paracetamol sulfate in the third compartment. The construction of a computer program will be left to problem #4.

It is usually the case that compartmental analysis of human physiological phenomena is more complex because such analysis requires many compartments and thus, many "paths" by which substances may be transferred. The most general three compartment model is shown in fig. 9.14.



General Three Compartment Model

Fig. 9.14

The equations governing the behavior of the system are obtained in the usual manner. They are:

```
Q1(I+1)=Q1(I)+(F1+T2*Q2(I)+K3*Q3(I)-(K1+E1+T1)*Q1(I))*D1

Q2(I+1)=Q2(I)+(F2+T3*Q3(I)+K1*Q1(I)-(K2+E2+T2)*Q2(I))*D1

Q3(I+1)=Q3(I)+(F3+K2*Q2(I)+T1*Q1(I)-(K3+E3+T3)*Q3(I))*D1
```

The quantities, F1, F2, and F3, denote the entering flows to each compartment. These flows are usually specified and are sometimes called the "drive" flows in analogy with the notion of driving forces in vibration theory. The quantities, E1, E2 and E3, are the respective transfer coefficients relating the concentrations in the compartments to exiting flows which do not enter other compartments. K1, K2, K3 and T1, T2, T3 are transfer coefficients. It is usually the case that T1=K1, T2=K2 and T3=K3. A principle use of such models is to obtain an understanding of the chemical kinetics of a physiological phenomena. It is frequently the case that more than one theoretical model may describe a system. Sprinson and Rittenberg (1949) describe a two compartment model for the description of the metabolism of organic nitrogen, and in 1951 Rittenberg constructed a three parameter model of the same phenomena. The decision as to which model is the most valid is a difficult decision and the fourth assumption is usually invoked when attempting to make the decision.

Probably the most difficult part of compartmental analysis is the determination of the transfer coefficients. The usual procedure for their determination rests on a comparison of numerical results with experimental results. A set of values for the transfer coefficients is assumed in the model and the numerical results obtained by using these coefficients is compared to the experimental results. If satisfactory agreement is not obtained, the coefficients are altered until the desired agreement is obtained. This is a process which is easy to describe but difficult to actually carry out and we will discuss it later in this chapter. Because of the difficulty of obtaining the constants of proportionality, compartmental models are usually restricted to a small number of compartments, even though the construction of computer programs for the analysis of models consisting of several compartments, is rather straightforward. The problem of determining the transfer coefficients is called the inverse problem, Monot and Martin (1974).

The Tracer Method

Since most of the experimental data used to determine the rate constants is obtained with the aid of tracer methods, it is appropria to give a short summary of the method. The texts by Sheppard (1962) and Atkins (1969) contain a more complete exposition of the subject.

It is exceedingly difficult to obtain, by direct means, the tran fer coefficients of substances in a biochemical reaction occurring i a living entity. These coefficients are usually obtained by an indirect method called the tracer method. This method consists in "labelling" or "tagging" the substance to be measured with a "tracer" which has the property that it is easily detected by an observer. Frequently, the tracer is a radioactive isotope; however, occasionally dyes ar also used as tracers. In essence, the tracer method consists in addi a small amount of the tracer to the substance of interest, called the mother substance. As the 'mother' substance is transferred from one co partment to another compartment in the system, a proportion of the tracer is also transferred as it is carried along with the mother substance. There is thus a change of tracer concentration in each compartment and the time variation of the tracer concentration in the compartments can be experimentally noted. Because it is assumed that the tracer does not alter the rate of transfer of the mother substance, the transfer coefficients used in describing the tracer time behavior are identical with the transfer coefficients used in describing the time behavior of the mother substance. A comparison of the experimental results with the numerical results obtained from a theoretical model of the transfer of the tracer permits the determination of the transfer coefficients. These coefficients may then be used to construct a model of the transfer of the mother substance and a computer based so tion of this model gives the dynamic behavior of the mother substance in the system. In this way the tracer method is an excellent example of the simultaneous use of theory and experiment to obtain understand ing.

Many biochemical processes are steady state processes and since their analysis requires the use of tracer methods, we present a brief discussion of the notion of a steady state. Before giving a formal definition of steady state, we give an example of its occurrence. Co sider the exchange of water between the plasma and the ascitic fluid.

It is known that water enters the plasma through the large intestine, leaves the plasma through the kidney, and that there is a flow of water from the plasma to the ascitic fluid and return. The two compartment configuration shown in figure 9.7 can be used to represent this system if the first compartment denotes the plasma and the second compartment denotes the ascitic fluid. Q1 and Q2 denote the amounts of water in each of the respective compartments. The entering flow of water to the plasma is denoted by F1 and the exiting flow of water to the kidney, F3, is described by the transfer coefficient, K3. The exchange rates of water between the plasma and the ascitic fluid are associated with the respective transfer coefficients, K1 and K2. The dynamic process of the exchange of water between the plasma compartment and the ascitic fluid compartment is said to be in a steady state process because there are actual flows of water yet there is no change in the quantity of water in each of the compartments. For this reason the quantities of interest are the transfer rates involved in the exchange process. The equation governing the quantity of water in the ascitic fluid can be written in the form

$$Q2(I+1) - Q2(I) = (K1*Q1(I) - K2*Q2(I))*D1.$$

Now, if it is assumed that there is no change in the amount of water in the ascitic fluid from one period to the next, it must be the case that

$$K1*Q1(I) = K2*Q2(I).$$

However, both K1 and K2 are constant and since Q2(I) is also assumed to be constant, it follows that Q1(I) must also be constant. Thus, the assumption of no change in the amount of water in the ascitic fluid from one period to the next implies that there is also no change in the amount of water in the plasma from one period to the next. There is then a steady state.

With this example as a background, we now define a steady state process. Our definition follows closely that given by Atkins (1969). A "Steady State Process", or "Steady State", is said to exist in a system of a mixture of substances if they are transported from one part of the system to another, or are transformed from one into anothe and yet, because their rates of removal are equal to their rates of replacement, their concentrations or amounts in all relevant compartments remain constant during the interval over which observations are made. If there are no flows into or out of the system, the system is called a "Closed System". A closed system which is in the steady state is said to be in a state of "Dynamic Equilibrium".

The importance of tracer techniques to the successful analysis of steady state phenomena is vividly pointed out by Riggs (1972) who states that, "Without the use of isotopic tracers it is always difficul and often impossible, to study the dynamics of such a system because no observable changes in the concentrations of S occur unless we deliberately add an appreciable amount of S to one of the compartments. But the moment we add S to one of the compartments, we destroy the steady state which is the very thing we want to study! It is the prime virtue of isotopic tracers, and particularly radioactive tracers, that they allow us to avoid this difficulty." The substance S referred to by Riggs in this quote is the same as our mother substance.

A further use of tracer kinetics is the determination of the rates of synthesis of products of biochemical reactions. This determination is possible because it is frequently the case that the substance being transported is created as a result of a biochemical reaction. Consequently, a determination of the rate at which the substance is being transported from one compartment to another provides a direct measure of the rate at which the substance is being synthesized. The work of Popjåk and Beeckmans (1950), in determining the rate of synthesis of cholesterol by growing fetuses is an example of the use of tracer methods in this manner.

As an example of the tracer method, we consider the determination of the amount of body water in a man. The patient receives a small, but known, amount of "labeled" water. The label is usually deuterium or a radioactive isotope such as tritium and its concentration in the water is also known. After sufficient time has elapsed, it is assumed that the labeled water has thoroughly mixed with the body fluids and a sample of body fluid is taken. Since the action of mixing with the body fluid will have diluted the concentration of the labeled substance in proportion to the amount of body fluid, a measurement of the concentration of the label in the sample will enable the determination of the mass of the body fluid. For example, suppose that 1000 cc of labeled water containing 120,000 counts per minute was injected. After mixing, a sample of 10 cc of body fluid is taken and is found to have an activity of 20 counts per minute. Since the original concentration of labeled water was 120 counts per minute per cc and the final concentration was 2 counts per minute per cc, there has been a dilution of the initial concentration by a factor of 1/60. Thus, the total amount of body fluid is 60*1000 cc or 60 liters.

This example is an illustration of "isotopic dilution", that is, the use of an isotope to determine the amount of a substance present in a system. The student should note the close analogy of the tracer method with the technique of tagging a small number of fish in a pond for the purpose of determining the total number of fish in the pond. In this technique, the number of fish that are tagged is noted and the tagged fish are then released into the lake. After a sufficient amount of time, the tagged fish are assumed to have thoroughly mixed with the other fish in the pond. A prescribed number of fish are again caught and the ratio of the number of tagged fish in this catch to the original number of tagged fish released into the pond enables an estimation of the total number of fish in the pond.

Intuitively, it is seen that a tracer should have the following properties:

(1) The biological or physiological system should not be able to distinguish between the mother substance and the tracer

(2) There should be no exchange of the tracer with other constituents in the system. This implies that the only change in the tracer concentration should be that due to the transporting of a portion of the tracer as it is carried along with the transferring mother substance. Thus, the transfer coefficients describing the dynamics of the tracer are the same as the transfer coefficients describing the exchanges of the mother substance.

(3) The tracer must be such that it is possible to accurately describe its variation in time by the use of a model. The model must reflect the transfer rates and the transformation of the mother substance in the system and in this way permit the determination of the time variation of the substance. The model may be expressed in a programming language or in the language of mathematics.

These properties imply that, when a tracer is added to the system, the kinetics of the system is not disturbed. This is usually accomplished by labelling the substance with a very small amount of the tracer.

Because the absolute radiation level of the tracer used in an experiment may be arbitrarily set by the investigator and because we are interested in the proportionate change in the tracer concentration it is convenient to formulate tracer kinetics problems in terms of fractional amounts or ratios. The student will recall that in the preceding example, it was the dilution, or proportionate decrease, of the original concentration of the tracer that was significant; not the change in absolute amounts of the tracer. For this reason, the term "specific activity" or "concentration" of the tracer is most helpful. The specific activity or concentration of a tracer is defined to be the ratio of the amount of the tracer to the amount of the mother substance. If the tracer is radioactive, the amount of radioactivity is measured in terms of counts per minute or a counting rate. Because the amount of the mother substance is measured in volume or mass units it is usual to measure the concentration, or specific activity, in counting rates per m-mole or counts per minute per milligram. This is due to the fact that the efficiency of the counting device is assumed to be constant but unknown. For this reason the term "relative specific activity" is useful. The relative specific activity of a tracer, or traced substance, is defined to be the ratio of the specific activity of the substance at a given time to the specific activity of the substance at a different time. It is usually the case that instruments for measuring radioactive tracer emissions measure the relative specific activity. Thus, the time variation of the concentration of the radioactive tracer is obtained by first determining, in some manner, the specific activity of the tracer in the mother substance at an initial time and then using the instrument to measure the relative specific activity at other instants of time with respect to the specific activity recorded at the initial time. The relative specific activity is also the ratio of the specific activity of one tracer substance, at a given time, to the specific activity of another tracer substance at the same time. This latter definition is particularly useful when comparing the relative activities of two or more tracers.

If the isotope is stable, its concentration is usually measured in terms of an "abundance ratio" which is defined to be the ratio of the number of atoms of the tracer isotope to the number of atoms of the most abundant natural isotope. A mass spectrometer is an instrument for detecting such a ratio. Since the number of atoms of the most abundant isotope is either known beforehand or is determinable, the time variation of the abundance ratio as measured by the mass spectrometer, enables the determination of the variation of the number of atoms or the time variation of the concentration of the stable tracer isotope.

With this brief summary of tracer methodology we introduce some notation that will be useful in constructing our computer models. Let:

$M1$ = Amount of the mother substance measured in grams, cc., etc.

$R1$ = Absolute amount of the tracer measured in counts per minute microcuries, etc.

$Q1$ = Amount of the labeling substance measured in grams, cc., et

$C1$ = Concentration of the tracer in the labeling substance measured in counts per minute per gram, microcuries per cc. etc.

$A1$ = Concentration of the tracer in the total amount of the labeled substance. It is measured in the same units as $C1$.

$C1$ is also called the "specific activity" of the original labeling substance and similarly $A1$ is often called the specific activity of the labeled mother substance. The total amount of the labeled substance, $T1$, is the sum of the amounts of the labeling substance and the amount of the mother substance. Thus,

$$T1 = Q1 + M1.$$

In terms of this notation it is seen that

$$C1 = R1/Q1$$

and

$$A1 = R1/(Q1 + M1).$$

Since the quantity of labeling substance added to the mother substanc is usually very much less than the amount of the mother substance, i.e. $Q1 \lll M1$, there is very little loss of accuracy if we write

$$A1 = R1/M1.$$

In fact, in many texts this equation is given as the definition of specific activity or concentration of the labelled substance.

To assist the student in understanding the notation we illustrate the application of the notation to the previous example of isotope dilution. In terms of this example, we have

$$Q1 = 1000 \text{ cc,}$$

$$R1 = 120{,}000 \text{ counts per min.}$$

and thus

$$C1 = 120 \text{ counts per min. per cc.}$$

Now

$$T1 = M1 + 1000$$

and

$$A1 = 120{,}000/(M1 + 1000).$$

From the sample it was determined that $A1 = 2$ counts per min. per cc. and thus,

$$2 = 120{,}000/(M1 + 1000)$$

or

$$M1 = 59{,}000 \text{ cc.}$$

Since $1000 << 59{,}000$, no significant error is made by ignoring the amount of the labeling substance, 1000 cc., in the denominator in the above equation. In this event, the amount of the body fluid is then

$$M1 = 60{,}000 \text{ cc.}$$

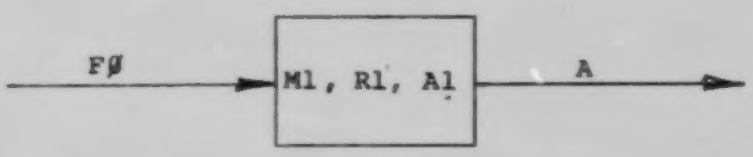
$$= 60 \text{ liters}$$

which agrees with our previous answer.

Determination of Transfer Coefficients

The problem of the determination of the transfer coefficients has frequently been mentioned in the preceding examples. The next example illustrates a method for obtaining transfer coefficients and will require an application of the search routines which were developed in an earlier chapter.

We consider the problem of determining the transfer coefficient occurring in a single compartment model of the turnover of inorganic phosphate in a rabbit plasma. It is known that inorganic phosphate enters the plasma of the rabbit and then leaves the plasma. A single compartment model is shown in figure 9.15.



Single Compartment Model of Transfer of
Plasma Phosphate in Rabbit Plasma
Fig. 9.15

The figure is a composite representation of the system and the notation appearing in the figure is to be interpreted in the following way. If the flow of the mother substance is to be considered, M1 is understood to represent the quantity of interest. On the other hand, if the flow of the ^{32}P-phosphate ion is the quantity of interest, the quantity R1 is to be understood. Finally, if the variation of the specific activity is the quantity of interest, it is understood that the symbol A1 is to be used. A single notation may be used to denote the exiting flow rate because the transfer coefficient is assumed to be the same for all three quantities. The entering flow rate of the mother substance will be denoted by F0(I).

The equation governing the behavior of the transport of the inorganic phosphate can be obtained by a proper accounting of the entering and exiting mass flows during an increment of time. In terms of the previous notation, the quantity of phosphate entering the compartment in a time period is

$$F0(I)*D1.$$

By Fick's law, the quantity of phosphate leaving the compartment in the same time period is

$$K1*M1(I)*D1,$$

and therefore the change in the amount of phosphate in the compartment during the time increment is

$$F0(I)*D1 - K1*M1(I)*D1.$$

An application of Fick's principle, which is a form of the fundamental law of change, gives

$$M1(I+1) = M1(I) +(F0(I) - K1*M1(I)*D1.$$

This is the equation governing the time behavior of the amount of inorganic phosphate in a rabbit plasma.

In order to use this equation, it is necessary to obtain the transfer coefficient, K1. The procedure for obtaining the transfer coefficient is the following:

(1) The inorganic phosphate is labeled with tracer ^{32}P-phosphate ions and a record of the time variation of the specific activity of radioactive ions is obtained.

(2) An equation is derived which describes the time variation of the specific activity of the tracer substance. Because the tracer is assumed to not alter the kinetics of the turnover of the inorganic phosphate, the equation will contain K1 as a parameter.

(3) The equation will be solved for various values of K1 and the value of the transfer coefficient which gives the least value for the measure of closeness of the numerical and the experimental results will be the desired value for the transfer coefficient.

The student should note that this procedure is entirely analogous to that used to obtain the growth coefficient in the Malthus model of the population growth of the United States. The value of K1 so obtained, can now be inserted into the equation governing the time variation of the inorganic phosphate and this equation solved to predict the variation in time of the phosphate. In this problem, and indeed in nearly all compartmental analysis problems, the structure of the equations governing the behavior of the mother substance and the behavior of the tracer substance are the same. Only the values of the parameters or the initial conditions may differ. This means that the solutions to the two sets of equations will be similar and hence the time variation of the mother substance and the tracer is also similar. Indeed, in many cases it is not necessary to refer again to the equation governing the mother substance since the information of interest can be obtained directly from the transfer coefficient or from an examination of the solutions of the equation governing the time variation of the specific activity of the tracer. For example, in some problems, the quantity of interest is the turnover time and this can be obtained from the transfer coefficient directly.

We now proceed to derive the equation governing the change of the labeling ^{32}P-phosphate. It is assumed that there is a thorough and immediate mixing of the tracer phosphate with the inorganic phosphate and hence the tracer ions are transported immediately along with the phosphate. The magnitude of the influx of the entering ^{32}P-phosphate is the product of the flow rate of the phosphate and the concentration of the ion, A0(I). Both F0(I) and A0(I) would have to be specified and entered as input data. In a single time increment, the quantity of ^{32}P-phosphate entering the compartment is

F0(I)*A0(I)*M1(I)*D1.

By the first assumption, the quantity of ^{32}P-phosphate leaving the compartment in the same period of time is

K1*R1(I)*D1

and thus, the change in the amount of ^{32}P-phosphate during the period is

F0(I)*A0(I)*M1(I)*D1 - K1*R1(I)*D1.

As before, an application of the fundamental law of change gives

$$R1(I+1) = R1(I) + (F0(I)*A0(I)*M1(I)-K1*R1(I))*D1. \qquad (9.15)$$

Now

$$R1(I) = A1(I)*M1(I), \qquad I=0,1,2,,,,$$

and thus the previous equation may be written as

$$A1(I+1)*M1(I+1) = A1(I)*M1(I)+((F0(I)*A0(I)-K1*A1(I)*M1(I))*D1. \qquad (9.1$$

If it is assumed that the system is in a steady state, that is

$$M1(I+1) = M1(I) = M1, \qquad I=0, 1, 2, , , ,$$

then the preceding equation becomes

$$A1(I+1) = A1(I) + (F0(I)*A0(I)/M1 - K1*A1(I))*D1. \qquad (9.17)$$

This equation, in conjunction with the experimentally obtained time variation of the ion, will be used to determine the rate constant.

As experimental data we use the data obtained by Hevesy and Hahn (1940) and presented in Atkins (1969). The data is shown in tabular form in table 9.1.

t		25	50	75	100	125	150	175	200	225	250
Q	0	.087	.11	.125	.13	.1407	.1407	.1375	.1405	.145	.151

Time Variation of Plasma Phosphate (From Atkins, 1969)

Table 9.1

Because Hevesey and Hahn used a constant rate of tracer infusion, the quantity $F0(I)*A0(I)/M1$ must also be constant. This constant value will be denoted by $F0$. In the experiment, it was assumed that the initial tracer concentration was zero. Thus,

$$A1(0) = 0.$$

Since the infusion rate of the tracer was not specified, it is necessary to determine $F0$ as well as the rate constant $K1$. For ease of discussion, we will use the single parameter determination program described in figure 4.2. In order to readily use the program, the rate constant, $K1$, will be denoted by A, and equation (9.17) will be written as

$$A1(I+1) = A1(I) + (F0 - A*A1(I))*D1. \qquad (9.18)$$

The closeness criteria will be the least squares criteria and hence the only significant change to the program is the alteration of the subroutine which calculates the measure of closeness, M1.

The computer program appears in figure 9.16 and the student should note that some of the headings, as well as the dimension statements, have been altered from those appearing in the original program. The change necessary to calculate M1 was accomplished by replacing lines 300 to 310 in the original program with lines 300 to 330 in the new program. Lines 300 to 330 are necessary to calculate the transport of the phosphate corresponding to a time increment of one minute. Thus, D1 = 1 minute. The calculated and experimental data are compared at 25 minute intervals. Since there are 10 such intervals, the index I has a range of 0 to 10 and the total time interval is 250 minutes. If a fundamental time interval differing from one minute is to be used, this section of the program must be altered in accordance with the change in the magnitude of D1. In addition, a change in the time increment will necessitate a change in the time unit used to express the rate constant. Lines 500 to 525 of the original program have also been changed to permit the entry of the empirically determined time variation of the specific activity, E(I), of the phosphate ion. Finally, lines 27 and 340 of the new program are necessary to insure the maintenance of the proper starting values each time the subroutine for the evaluation of M1 is used.

The results of a typical run are listed in figure 9.17. The first column listed under the heading, "The values of A, M1 and H are" displays the successive values of the rate constant as the program searches for the best value of the rate constant. The second column displays the value of the closeness criteria, M1, and the third column lists the search step size. In this example, A has been determined to four significant digits. Because the closeness criteria is the least squares criteria, the best value for A is that value which minimizes the sum of the squares of the deviations between the specific activities as calculated with a particular value of the rate constant and the specific activities as empirically determined. An examination of the program will reveal that as F0, the constant infusion rate, is varied, the magnitude of the transfer coefficient also varies. Thus, as stated before, this is really a two parameter problem. The situation is very similar to the constant environment population model comparison with the United States population data.

```
1 REM   DETERMINATION OF TRANSFER COEFF. IN ONE COMPARTMENT MODEL
5 REM  DATA FROM HEVESY AND HAHN,  ATKINS(1969)
8 REM   D1,  THE TIME INCREMENT IS ONE MINUTE
10 DIM A1(30),A2(30),E(30)
12 PRINT "TYPE THE INFLUX RATE F0"
13 INPUT F0
14 PRINT
15 GOSUB 500
18 REM   A DENOTES THE TRANSFER COEFFICIENT
20 PRINT "TYPE THE INITIAL GUESS A, AND THE INIT. CONCENTRATION A1(0)"
25 INPUT A,A1(0)
26 PRINT
30 PRINT "INPUT THE INIT. STEP SIZE  H  AND THE LIM. STEP SIZE H1"
35 INPUT H,H1
36 PRINT
40 PRINT "INPUT THE MAX. NO. OF ALLOWABLE STEPS,  C1"
45 INPUT C1
46 PRINT
47 PRINT
48 PRINT "THE VALUES OF  A,  M1  AND  H  ARE"
49 PRINT
50 LET C=0
55 GOSUB 300
60 LET M0=M1
65 LET A=A+H
70 GOSUB 300
95 IF M1<=M0GO TO 110
100 LET A=A-H
105 GO TO 200
110 LET C=C+1
115 IF C<C1GO TO 125
120 GO TO 400
125 LET A=A+H\LET M0=M1
130 GOSUB 300
135 IF M1<=M0GO TO 110
137 LET A=A-H
140 LET H=H/10
145 IF H<=H1GO TO 445
150 GO TO 65
200 LET A=A-H
205 GOSUB 300
210 IF M1<=M0GO TO 225
215 LET A=A+H
220 GO TO 140
225 LET C=C+1
230 IF C<=C1GO TO 240
235 GO TO 400
240 LET M0=M1
245 GO TO 200
```

Inverse Problem for a Single Compartment Model

Fig. 9.16

```
295 REM   INSTR. NOS.  300  TO  385  EVALUATE M1
296 REM   A1(J)  DENOTES QUANTITIES CALCULATED USING  D1=1  MINUTE
297 REM   A2(I)  DENOTES THE VALUE OF  A1(J)  AT MULTIPLES OF 25 MINUTES
298 REM   THE  J  INDEX COUNTS THE ONE MINUTE INTERVALS
299 REM   THE  I  INDEX  COUNTS THE  25  MINUTE INTERVALS
300 LET A1(0)=A3
301 LET A2(0)=A1(0)
302 LET D1=1
305 FOR I=0 TO 9
307 FOR J=0 TO 24
310 LET A1(J+1)=A1(J)+(F0-A*A1(J))*D1
315 NEXT J
320 LET A2(I+1)=A1(J+1)
325 LET A1(0)=A2(I+1)
330 NEXT I
340 LET A1(0)=A3
350 LET S=0
355 FOR I=0 TO 10
360 LET D=ABS(A2(I)-E(I))
365 LET S=S+D*D
370 NEXT I
375 LET M1=S
380 PRINT A,M1,H
385 RETURN
400 PRINT "EXCEEDED MAX. NO. OF STEPS"
405 PRINT "THE VALUES OF  A  AND  M0  ARE"
410 PRINT A,M0
415 GO TO 460
445 PRINT
446 PRINT
450 PRINT "SEARCH COMPLETE.  THE VALUES OF A,  M0  AND  C  ARE"
455 PRINT A,M0,C
459 PRINT
460 PRINT
462 REM  LINES 460-485 PRINT THE BEST VALUES OF A2(I) AND E(I)
465 PRINT " I                    A2(I)             E(I)"
470 PRINT
475 FOR I=0 TO 9
480 PRINT I,A2(I),E(I)
485 NEXT I
490 GO TO 540
495 REM    INSTR. NOS. 500 TO 525 ENTER THE EXPERIMENTAL DATA
500 DATA 0,.087,.11,.125,.13,.1407,.1407
505 DATA .1375,.1405,.145,.151
515 FOR J=0 TO 10
520 READ E(J)
525 NEXT J
530 RETURN
540 END
```

Fig. 9.16 (Cont.)

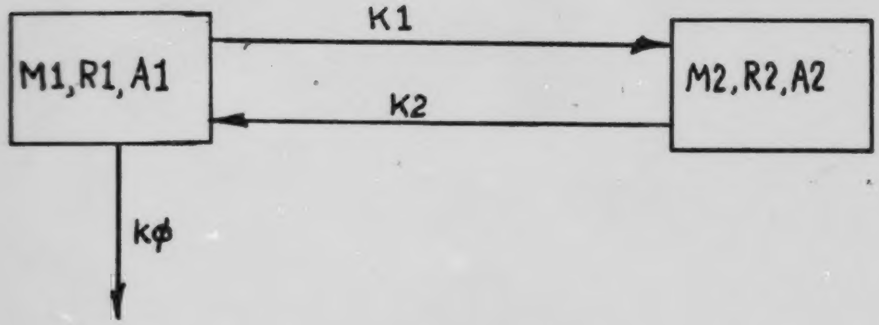
```
TYPE THE INFLUX RATE F0
?.005

TYPE THE INITIAL GUESS A, AND THE INIT. CONCENTRATION A1(0)
?.035, 0

INPUT THE INIT. STEP SIZE  H  AND THE LIM. STEP SIZE H1
?.001, .00001

INPUT THE MAX. NO. OF ALLOWABLE STEPS,  C1
?100


THE VALUES OF  A,  M1  AND  H  ARE

 .035          3.31007E-04   1.00000E-03
 .036          3.16687E-04   1.00000E-03
 .037.         5.18786E-04   1.00000E-03
 .0361         3.27697E-04   1.00000E-04
 .0359         3.07838E-04   1.00000E-04
 .0358         3.01181E-04   1.00000E-04
 .0357         2.96752E-04   1.00000E-04
 .0356         2.94583E-04   1.00000E-04
 .0355         2.94709E-04   1.00000E-04


SEARCH COMPLETE.  THE VALUES OF A,  M0  AND  C  ARE
 .0356         2.94583E-04   5


 I               A2(I)         E(I)

 0             0             0
 1             .0837896      .087
 2             .117732       .11
 3             .131482       .125
 4             .137052       .13
 5             .139309       .1407
 6             .140223       .1407
 7             .140593       .1375
 8             .140743       .1405
 9             .140804       .145
```

Output from Program Shown in Fig. 9.16

Fig. 9.17

By actually varying F0 and comparing the results obtained with each variation, your author found a reasonably close agreement (in the sense of the sums of the squares of the deviations) with the experimental data. The student should modify the two-dimensional search routine, figure 5.5, and attempt to obtain the "best" value for both F0 and K1.

As a second example, we consider the problem of the determination of the rate constants in a two compartment model used to explain the exchange of water between the plasma and the extracellular fluid. It is also assumed that water is excreted from the plasma as urine. Insuline was used as a tracer. The example is taken from Atkins (1969), and the system model is portrayed in figure 9.18.



Two Compartment Model for Exchange of Water Between Plasma and Extracellular Fluid

Fig. 9.18

The first compartment represents the plasma and the second compartment represents the extracellular fluid. The experiment was performed by Zender, Denkinger and Falbriard (1965) who injected insulin intraveneously and recorded the time variation of the insulin in both the plasma and the extracellular fluid.

The equations describing the diffusion of the insulin are derived in a manner analogous to that used in the previous example. Let R1(I) and R2(I) denote the amounts of insulin in compartments 1 and 2 respectively. The amount of insulin entering the first compartment in a time period is

$$K2*R2(I)*D1$$

and the amount of insulin leaving the same compartment in the same time period is

$$K1*R1(I)*D1 + K0*R1(I)*D1 .$$

This expression results from the first assumption and the designation of the transfer coefficients by K0 and K1. An application of the fundamental law of change gives

$$R1(I+1) = R1(I)+K2*R2(I)*D1-K1*R1(I)*D1-K0*R1(I)*D1.$$

The equation governing the change of the insulin in the second compartment is derived in a similar manner. It is

$$R2(I+1) = R2(I)+K1*R1(I)*D1-K2*R2(I)*D1.$$

Because

$$R1(I) = A1(I)*M1(I) \quad \text{and} \quad R2(I) = A2(I)*M2(I) \quad I=0,1,2,,,$$

the above equations may be written as

$$A1(I+1)*M1(I+1) = A1(I)*M1(I)+K2*A2(I)*M2(I)*D1-K1*A1(I)*M1(I)*D1$$
$$-K0*A1(I)*M1(I)*D1$$

and

$$A2(I+1)*M2(I+1) = A2(I)*M2(I)+K1*A1(I)*M1(I)*D1-K2*A2(I)*M2(I)*D1.$$

Since the system is assumed to be in a steady state, the quantities of water in each compartment, M1(I) and M2(I), are constant. Thus, we can write

$$M1(I) = M1 \quad \text{and} \quad M2(I) = M2.$$

Furthermore, the assumption of a steady state implies that the rate constants K1 and K2 are equal. These assumptions enable the preceding equations to be written in the simpler form:

$$A1(I+1) = A1(I)+K1*A2(I)*(M2/M1)*D1-K1*A1(I)*D1$$

$$-K0*A1(I)*D1$$

and

$$A2(I+1) = A2(I)+K1*A1(I)*(M1/M2)*D1 - K1*A2(I)*D1.$$

In most mammals the ratio of the amount of water in the plasma to the amount of water in the extracellular fluid is approximately four. Thus. we set

$$M2/M1 = 4$$

and get

$$A1(I+1) = A1(I)+(4*K1*A2(I)-K1*A1(I)-K0*A1(I))*D1$$

and

$$A2(I+1) = A2(I)+(K1*A1(I)/4 - K1*A2(I))*D1.$$

This is the desired form of the equations.

The data used for the determination of the rate constants K0 and K1 is taken from that given in Atkins (1969). Because there are two parameters to be obtained, the two-dimensional search program listed in figure 5.5 for the determination of the growth coefficie in the finite growth model will be used. The program will be modifie to accommodate the above two equations. The concentration of insuli in the first compartment has been empirically determined for 14 successive ten-minute intervals. This data constitutes the empirica data used in comparison with the calculated specific activities to select the "best" set of rate constants.

The following is a summary of the changes made in the original search program of figure 5.5.

(1) Lines 1 and 5 are remark statements describing the program.

(2) Line 8 is a modified dimension statement to accommodate the specific activities.

(3) Lines 12 and 14 are new input statements as are lines 22 and 24.

(4) The student will recall that in the original program, the fundamental search variables were labeled A and B. Lines 20 and 21 serve to identify the rate constants K0 and K1 with the variables A and B respectively.

(5) Lines 50 to 58 enter the empirically determined specific activities of the plasma. They are denoted by E1(I).

(6) Lines 26 and 27, in conjunction with lines 732 and 733 are necessary to insure the correct initial concentrations after a set of concentrations has been calculated. After a set of concentrations has been obtained and the corresponding measure of closeness evaluated, the specific activities A1(0) and A2(0) must be reset to their original starting values A5 and A6 respectively. This resetting is necessary because, at the end of a calculation of a complete set of specific activities for ten minute intervals from 0 to 140 minutes, the values of A1(0) and A2(0) are equal to the values of the specific activities corresponding to 130 minutes. This is due to the way the specific activities are calculated in instructions 705 to 730. By resetting A1(0) and A2(0) to their initial or starting values we insure that there is no accumulation of specific activities from a calculation of one set of values of the search variables to the calculation of the next set. This is a form of initializing.

(7) Lines 700 and 701 also initialize the specific activities.

(8) Lines 702 to 730 calculate the specific activities using a time increment of one minute. These instructions also save the specific activities corresponding to ten minute intervals. The latter pair of specific activities are labeled A3(I) and A4(I) respectively. The variables A1(J) and A2(J) denote the concentrations of the insulin in the first and second compartment corresponding to one minute intervals. The procedure to obtain the insulin values at ten minute intervals is to calculate the concentrations using a one minute interval and to then store every tenth value so obtained. The student should "walk through" this part of the program to insure himself that this is indeed accomplished.

(9) Lines 735 to 743 calculate the sum of the squares of the deviations corresponding to a given set of values of the rate constants.

(10) Lines 837 to 844 print out the results in order that a visual comparison can be made of the closeness of the calculated results to the empirical results.

If a "good" set of starting values is not available for the rate constants, the program can be used as an aid in finding them. This may be done by choosing the number of search steps, T, equal to one When the program is then run, the program will not attempt a search, but will calculate the measure of closeness, as well as print out the calculated concentrations for the guessed pair of rate constants. After a few trials, it is usually possible to find a pair of "reasonable" starting values. Using these values, a larger value for T, say 50 or 100, may then be used and the program will search or "try to refine" these initial values. If the search is successful, that is, if a set of rate constants are found which do minimize the sum of the squares of the deviations, then the values should be check to assure that a local minimum has not been obtained. The student should recall the methods for checking the results obtained from search routines as stated in chapter V. The modified program appears in figure 9.19 and results of a typical run are given in figures 9.20 and 9.21.

```
RBC60       BASIC V01-05

1 REM   DET. OF TRANS. COEFS., 2 COMPTS., 2 VAR. SEARCH
5 REM   D1, THE TIME INCREMENT IS ONE MINUTE
8 DIM A1(20),A2(20),A3(20),A4(20),E1(20)
12 PRINT "TYPE THE INITIAL GUESSES FOR THE RATE CONSTANTS, K0 AND K1"
14 INPUT K0,K1
20 LET A=K0
21 LET B=K1
22 PRINT "TYPE THE INITIAL CONCS.  A1(0)  AND  A2(0)"
24 INPUT A1(0),A2(0)
26 LET A5=A1(0)
27 LET A6=A2(0)
28 PRINT
30 PRINT "H  AND  K  ARE THE INITIAL STEP SIZES"
31 PRINT "E1  AND  E2  ARE THE LIMITING STEP SIZES"
32 PRINT "INPUT  H,  K,.  E1,  E2"
33 INPUT H,K,E1,E2
34 PRINT
37 PRINT "T IS THE MAXIMUM ALLOWABLE NO. OF SEARCH STEPS"
38 PRINT "INPUT T"
39 INPUT T
40 PRINT
42 PRINT "A1  AND  A2  ARE THE MIN. AND MAX. PTS. OF  A  INTERVAL"
43 PRINT "B1  AND  B2  ARE THE MIN. AND MAX. PTS. OF  B  INTERVAL"
44 PRINT "INPUT A1, A2, B1 AND B2"
46 INPUT A1,A2,B1,B2
48 REM  INST. NOS. 50 TO 70 ARE DATA INPUT
50 DATA 3350,1688,944,593,415,314,250,204
52 DATA 170,142,119,100,85,71,60
54 FOR J=0 TO 14
56 READ E1(J)
58 NEXT J
84 REM INITIALIZING
86 LET C=0
93 PRINT
94 PRINT
95 PRINT "THE VALUES OF  A,  B,  M0  AND  C  ARE"
100 GOSUB 700
105 LET M0=M1
110 LET A=A+H\LET B=B
120 GOSUB 700
135 IF M1<M0GO TO 145
140 LET A=A-H\LET B=B
142 GO TO 250
145 LET A=A+H\LET B=B\LET M0=M1
160 GOSUB 700
175 IF M1<M0GO TO 185
180 LET A=A-H\LET B=B
182 GO TO 400
185 IF A<A2GO TO 145
195 LET A=A\LET B=B\LET M0=M1
200 PRINT A,B,M0
205 PRINT "EXCEEDED ALLOWED MAX. VALUE OF  A"
206 PRINT "THE VALUES OF  A,  B  AND  M0  ARE"
207 PRINT A,B,M0
208 STOP
210 STOP
```

522

Fig. 9.19

```
250 LET A=A-H\LET B=B
255 GOSUB 700
270 IF M1<M0GO TO 280
275 LET A=A+H\LET B=B
277 GO TO 400
280 LET A=A-H\LET B=B\LET M0=M1
285 GOSUB 700
300 IF M1<M0GO TO 310
305 LET A=A+H\LET B=B
307 GO TO 400
310 IF A>A1GO TO 280
320 LET A=A\LET B=B\LET M0=M1
325 PRINT A,B,M0
330 PRINT "EXCEEDED ALLOWED MIN.  VALUE OF  A"
331 PRINT "THE VALUES OF  A,  B  AND  M0  ARE"
332 PRINT A,B,M0
335 STOP
350 PRINT
400 LET A=A\LET B=B+K
405 GOSUB 700
420 IF M1<M0GO TO 430
425 LET A=A\LET B=B-K
427 GO TO 500
430 IF B<B2GO TO 445
435 LET A=A\LET B=B\LET M0=M1
438 PRINT "EXCEEDED ALLOWED MAX.  VALUE OF  B"
439 PRINT "THE VALUES OF  A,  B  AND  M0  ARE"
440 PRINT A,B,M0
444 STOP
445 LET A=A\LET B=B+K\LET M0=M1
450 GOSUB 700
465 IF M1<M0GO TO 430
470 LET A=A\LET B=B-K
472 GO TO 100
475 PRINT
500 LET A=A\LET B=B-K
505 GOSUB 700
520 IF M1<M0GO TO 528
525 GO TO 680
528 IF B>B1GO TO 540
530 LET A=A\LET B=B\LET M0=M1
535 PRINT "EXCEEDED ALLOWED MIN.  VALUE OF  B"
536 PRINT "THE VALUES OF  A,  B  AND  M0  ARE"
537 PRINT A,B,M0
538 STOP
540 LET A=A\LET B=B-K\LET M0=M1
545 GOSUB 700
560 IF M1<M0GO TO 528
565 LET A=A\LET B=B+K
567 GO TO 100
570 PRINT
620 PRINT
645 PRINT
650 LET H=H/10\LET K=K/10
652 IF H<E1GO TO 658
653 IF K<E2GO TO 655
654 GO TO 100
655 LET K=10*K
656 GO TO 100
658 IF K<E2GO TO 662
659 LET H=10*H
660 GO TO 100
```

Fig. 9.10 continued

```
662 PRINT
663 PRINT
664 PRINT
665 PRINT "THE FINAL  VALUES OF  A,  B,  M0  AND  C  ARE"
667 PRINT A,B,M0,C
668 PRINT
669 PRINT "THE FINAL VALUES OF  H  AND  K  ARE"
670 PRINT H,K
675 GO TO 835
677 PRINT
680 LET A=A\LET B=B+K\LET M0=M0
681 PRINT
682 PRINT "THE INTERMEDIATE VALUES OF A, B, M0 ARE"
684 PRINT A,B,M0
686 GO TO 650
689 PRINT
690 REM INST. NOS. 700 TO 716 EVALUATE M
700 LET A1(0)=A5\LET A2(0)=A6
701 LET A3(0)=A5\LET A4(0)=A6
702 LET D1=1
705 FOR I=0 TO 14
707 FOR J=0 TO 9
710 LET A1(J+1)=A1(J)+(4*B*A2(J)-B*A1(J)-A*A1(J))*D1
712 LET A2(J+1)=A2(J)+(B*A1(J)/4-B*A2(J))*D1
715 NEXT J
720 LET A3(I+1)=A1(J+1)
722 LET A4(I+1)=A2(J+1)
724 LET A1(0)=A3(I+1)
726 LET A2(0)=A4(I+1)
730 NEXT I
732 LET A1(0)=A5
733 LET A2(0)=A6
735 LET S1=0
737 FOR I=0 TO 14
739 LET H1=E1(I)-A3(I)
741 LET S1=S1+H1*H1
743 NEXT I
815 LET M1=S1
817 PRINT
818 PRINT A,B,M1,C
819 REM  INST. NO 821 TO 830 PREVENT ENDLESS LOOPING
820 LET C=C+1
821 IF C<TGO TO 832
822 LET M0=M1
823 PRINT "THE LOST VALUES OF A,  B,  M0  AND  C  ARE"
824 PRINT A,B,M0,C
825 GO TO 835
827 PRINT "THE VALUES OF  H  AND  K  ARE"
828 PRINT H,K
829 PRINT "EXCESSIVE NUMBER OF STEPS"
830 STOP
832 RETURN
835 PRINT
836 REM  LINES 837 TO 844 PERMIT COMPARISON OF RESULTS
837 PRINT " COMPARISON OF EXPERIMENTAL AND CALCULATED VALUES"
838 PRINT
839 PRINT " I           E1(I)            A3(I)            A4(I)"
840 PRINT
841 FOR I=0 TO 14
842 PRINT
843 PRINT I,E1(I),A3(I),A4(I)
844 NEXT I
850 END
```

Fig. 9.19 continued

The portion of the output listed in figure 9.20 preceding the statement "The Values of A, B, M0 and C are" presents the computer request for input, together with your author's response. The remaining portion of the output lists results obtained from the run. The four columns appearing under the heading, "Comparison of Experimental and Calculated Values" in fig. 9.21, depict in order from left to right, the number of the ten-minute time interval, the experimental data corresponding to the time interval, the calculated concentrations of insulin in the plasma, and the calculated concentrations of the insulin in the extracellular fluid corresponding to the respective time interval.

```
TYPE THE INITIAL GUESSES FOR THE RATE CONSTANTS, K0 AND K1
?.04, .02
TYPE THE INITIAL CONCS.  A1(0)  AND  A2(0)
?3350, 0

H  AND  K  ARE THE INITIAL STEP SIZES
E1  AND  E2  ARE THE LIMITING STEP SIZES
INPUT  H,  K,  E1,  E2
? .01, .01,  .001,  .001

T IS THE MAXIMUM ALLOWABLE NO. OF SEARCH STEPS
INPUT T
?100

A1  AND  A2  ARE THE MIN. AND MAX  PTS.  OF  A  INTERVAL
B1  AND  B2  ARE THE MIN.  AND MAX.  PTS.  OF  B  INTERVAL
INPUT A1, A2, B1 AND B2
?0, 1, 0, 1


THE VALUES OF  A,  B,  M0  AND  C  ARE

 .04              .02              68416.3        0

 .05              .02              28846.6        1

 .06              .02              215953         2

 .05              .03              48023.5        3

 .05              .01              79571.8        4

THE INTERMEDIATE VALUES OF A, B, M0 ARE
 .05              .02              28846.6

 .05              .02              28846.6        5

 .051             .02              40477.7        6

 .049             .02              19344.7        7

 .048             .02              12165.2        8

 .047             .02              7519.64        9

 .046             .02              5638.69        10

 .045             .02              6775.22        11

 .046             .021             3670.07        12

 .046             .022             2544.42        13

 .046             .023             2183.6         14

 .046             .024             2516.01        15
```

Fig. 9.20

```
 046         .023          2182.6        16

 047         .023          5291.93       17

 045         .023          2021.45       18

 044         .023          5077.7        19

 045         .024          2039.91       20

 045         .022          2751.42       21
THE INTERMEDIATE VALUES OF A, B, M0 ARE
.045         .023          2021.45


THE FINAL VALUES OF  A,  B,  M0  AND  C  ARE
 045         .023          2021.45        22

THE FINAL VALUES OF  H  AND  K  ARE
 1.00000E-04   1.00000E-04
```

Fig. 9.20 continued

```
COMPARISON OF EXPERIMENTAL AND CALCULATED VALUES
I           E1(I)            A3(I)            A4(I)

0              2350            2350            0
1              1688            1722.23         123.912
2              944             958.796         164.91
3              593             590.531         170.155
4              415             404.301         160.818
5              314             303.034         146.303
6              250             242.38          130.702
7              204             201.963         115.715
8              170             172.316         101.982
9              142             148.945         89.6575
10             119             129.642         78.7453
11             100             113.254         69.1107
12             85              99.1269         60.6354
13             71              86.848          53.1906
14             60              76.1292         46.6559
```

Fig. 9.21

Just as a matter of interest, your author compared the results obtained from this program with those obtained when the measure of closeness was chosen to be the sum of the squares of the relative errors. To obtain the rate constants corresponding to this measure of closeness, it was only necessary to alter line 739 to read

```
739  LET H1 = (A3(I) - E1(I))/E1(I)
```

and to then proceed as before. This single and simple change, which enables the determination of the parameters using an entirely different closeness criteria, again illustrates the versatility of the direct computer language approach.

The values of the rate constants corresponding to a minimum value of the sum of squares of the deviations were used as initial values for the altered program. The results corresponding to the closeness criteria of the sum of the squares of the relative errors are displayed in figure 9.22. The final values of the rate constants were achieved by setting T equal to a large number and carrying out the search. The results displayed in this figure were obtained by the simple device of setting the initial choices of the rate constants equal to thos obtained from the previous run and then setting T=1 so that only a summary would be printed out.

To facilitate discussion of the results, it is convenient to denote the expression, "sum of the squares of the deviations" by SSD and to denote the expression, "sum of the squares of the relative errors" by SSRE. A comparison of the results obtained from the two programs reveals that:

(1) For the SSD criteria, K0=0.0454 and K1=0.0233, whereas for the SSRE criteria K0=0.0497 and K1=0.0301. This is a difference of approximately 10% in K0 and 30% in K1.

(2) There is a very large difference in the numerical value of each measure of closeness. This is to be expected because the SSRE measure is a relative error and for each data point this measure should be less than unity in magnitude if the "fit is any good at all". In contrast, since the empirical values for the concentrations are very large, i.e. very much greater than unity in magnitude, it is to

```
RBC60

TYPE THE INITIAL GUESSES FOR THE RATE CONSTANTS, K0 AND K1
?.04970, .03011
TYPE THE INITIAL CONCS.  A1(0)  AND  A2(0)
?3350, 0

H  AND  K  ARE THE INITIAL STEP SIZES
E1  AND  E2  ARE THE LIMITING STEP SIZES
INPUT  H,  K,  E1,  E2
?.00001, .00001, .000001, .000001,

T IS THE MAXIMUM ALLOWABLE NO. OF SEARCH STEPS
INPUT T
?1

A1  AND  A2  ARE THE MIN.  AND MAX.  PTS.  OF  A  INTERVAL
B1  AND  B2  ARE THE MIN.  AND MAX.  PTS.  OF  B  INTERVAL
INPUT A1, A2, B1 AND B2
?0, 1, 0, 1


THE VALUES OF  A,  B,  M0  AND  C  ARE

 .0497          .03011         .0713295       0
THE LOST VALUES OF A,  B,  M0  AND  C  ARE
 .0497          .03011         .0713295       1

 COMPARISON OF EXPERIMENTAL AND CALCULATED VALUES

 I          E1(I)          A3(I)          A4(I)


 0             3350           3350           0

 1             1688           1540.01        154.733

 2             944            822.299        189.45

 3             593            518.022        182.847

 4             415            373.265        163.743

 5             314            292.601        142.449

 6             250            239.783        122.441

 7             204            200.716        104.702

 8             170            169.647        89.3324

 9             142            144.006        76.1451

 10            119            122.473        64.8769

 11            100            104.247        55.266

 12            85             88.7655        47.075

 13            71             75.5953        40.0965

 14            60             64.3837        34.1521
```

Results From SSRE

Figure 9.22

be expected that even if the fit is good, the deviations will also be greater than unity. Thus, the SSD criteria is a sum of squares of quantities which themselves are each larger in magnitude than unity, and the value of this criteri is therefore, much larger than the value of the SSRE criteria which is a sum of squares of quantities which should each less than unity in magnitude.

(3) The agreement between the calculated and empirical values for early times for the SSD criteria is quite good. Howeve as time increases, the agreement becomes worse. This is due to the fact that the magnitudes of the concentrations for early times are very much larger than the magnitudes of the concentrations for later times. This kind of agreem illustrates the well known fact that parameters determined with the aid of the SSD criteria tend to be of such a value that relatively close agreement is obtained for large calculated and empirical values but poor relative agreement is obtained for small calculated and empirical values. Not that when using the SSD criteria the relative error for the first time period is approximately 1% whereas for the last time period it is 25%. In contrast, the relative agreement between the calculated and the empirical values as obtained using the SSRE criteria is fairly constant for all data points. However, the magnitude the difference Al(1)-El(1) is 148 and this is very much larger than the corresponding magnitude of 19 using the SSD criteria for closeness. However, the agreement between Al(1) and El(1) measured in terms of per cent error is 8.4% and the agreement between the last values, Al(14) and El(14), when measured in terms of per cent error is 7%. This illustrates the fact that the SSRE criteria produces parameter values which correspond to calculated values which are in close agreement with the empirical results measured in terms of per cent error for all of the data points. There are exceptions to these statements; nevertheless your author felt it prudent to mention them because they are a part of the "lore" of curve fitting and parameter determination. Figure 9.23 is a comparison of the deviations, and figure 9.24 is a comparison of the rel

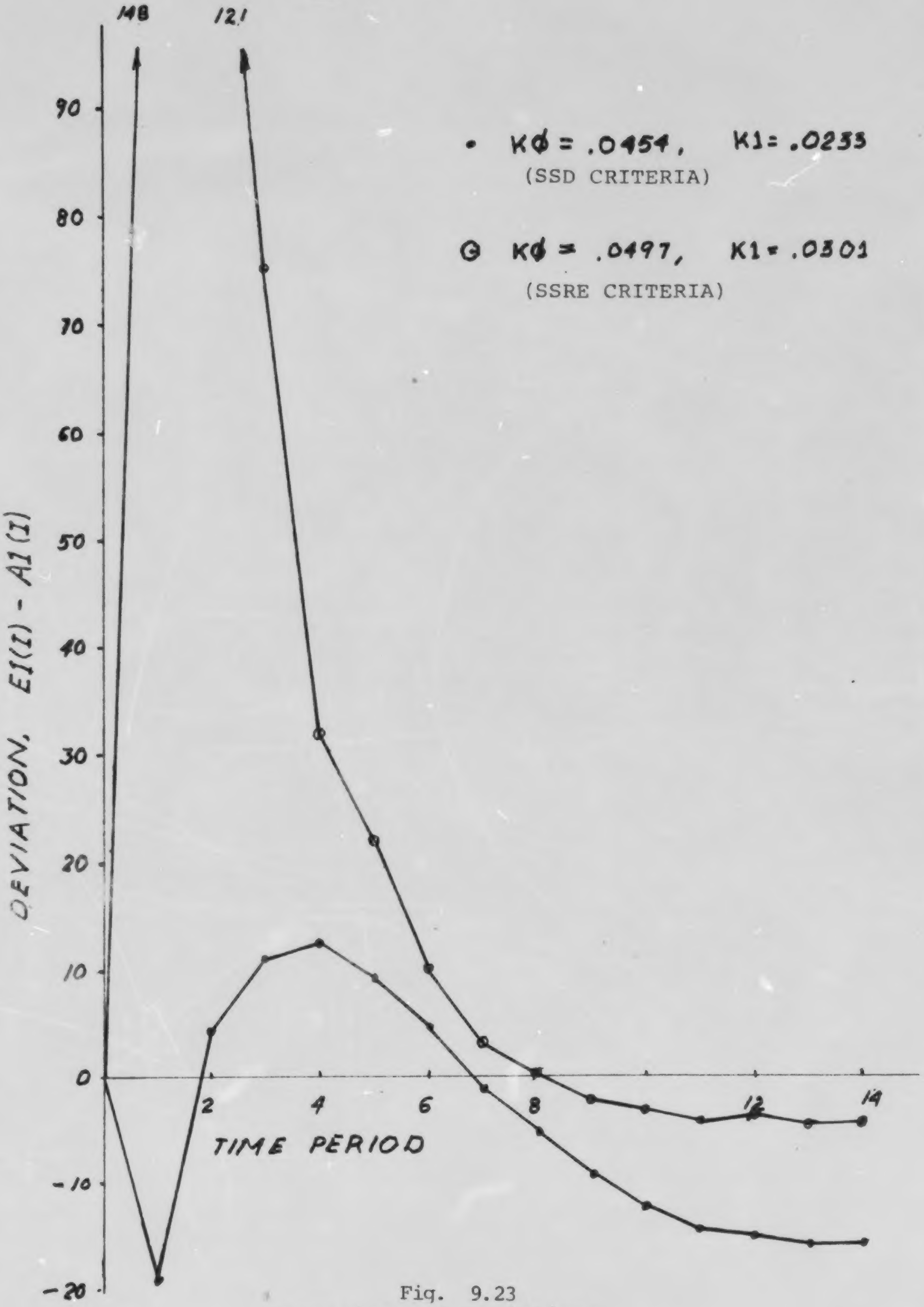


Fig. 9.23
Comparison of Deviations

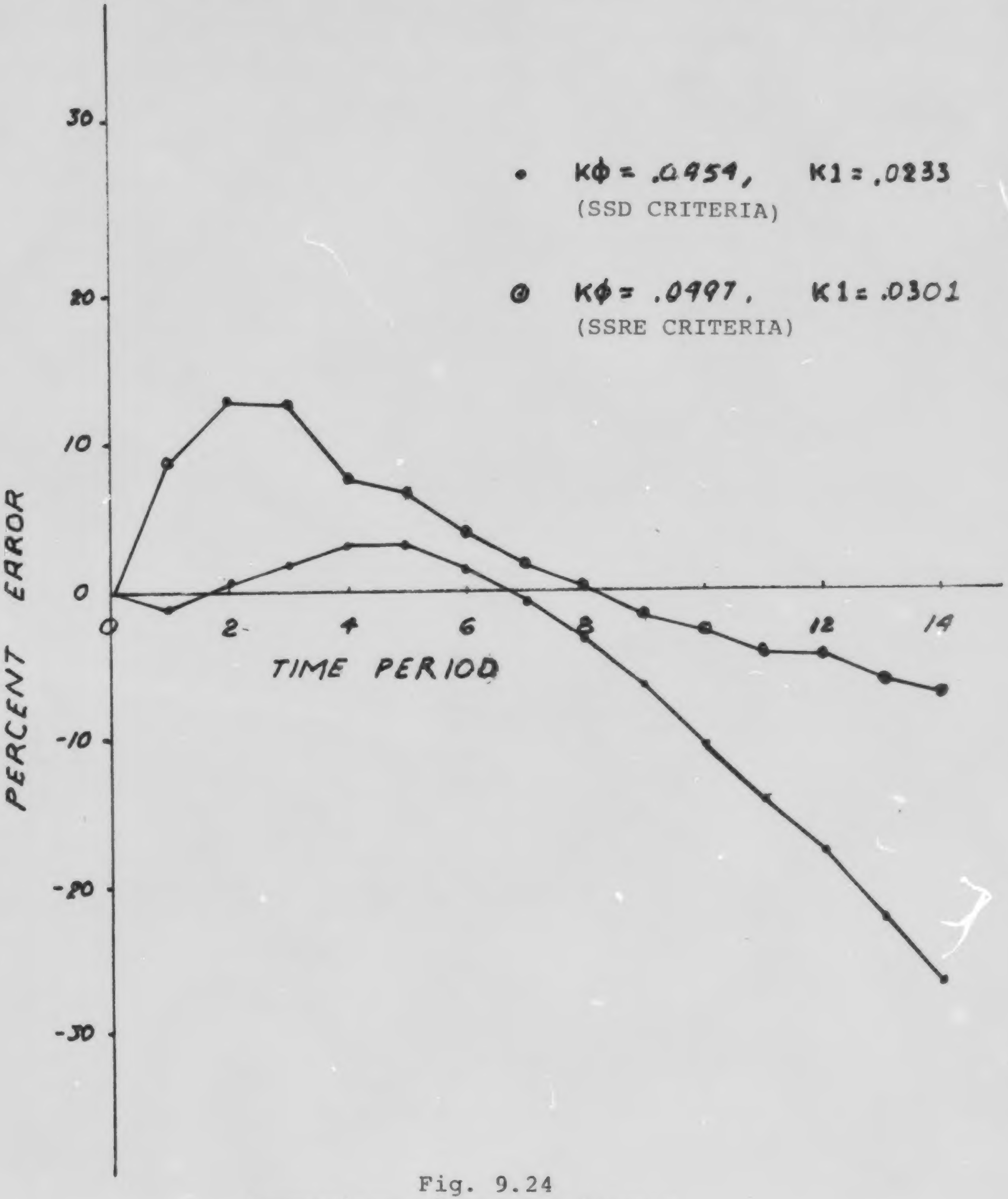


Fig. 9.24
Comparison of Percent Errors

tive errors obtained by using each criteria.

(4) There are considerable differences in the corresponding specific activities of the extracellular fluid, A2(I), as calculated using the SSD criteria and as calculated using the SSRE criteria. The values of maximum magnitudes or peaks are different and they occur at different times. A comparison is shown in figure 9.25.

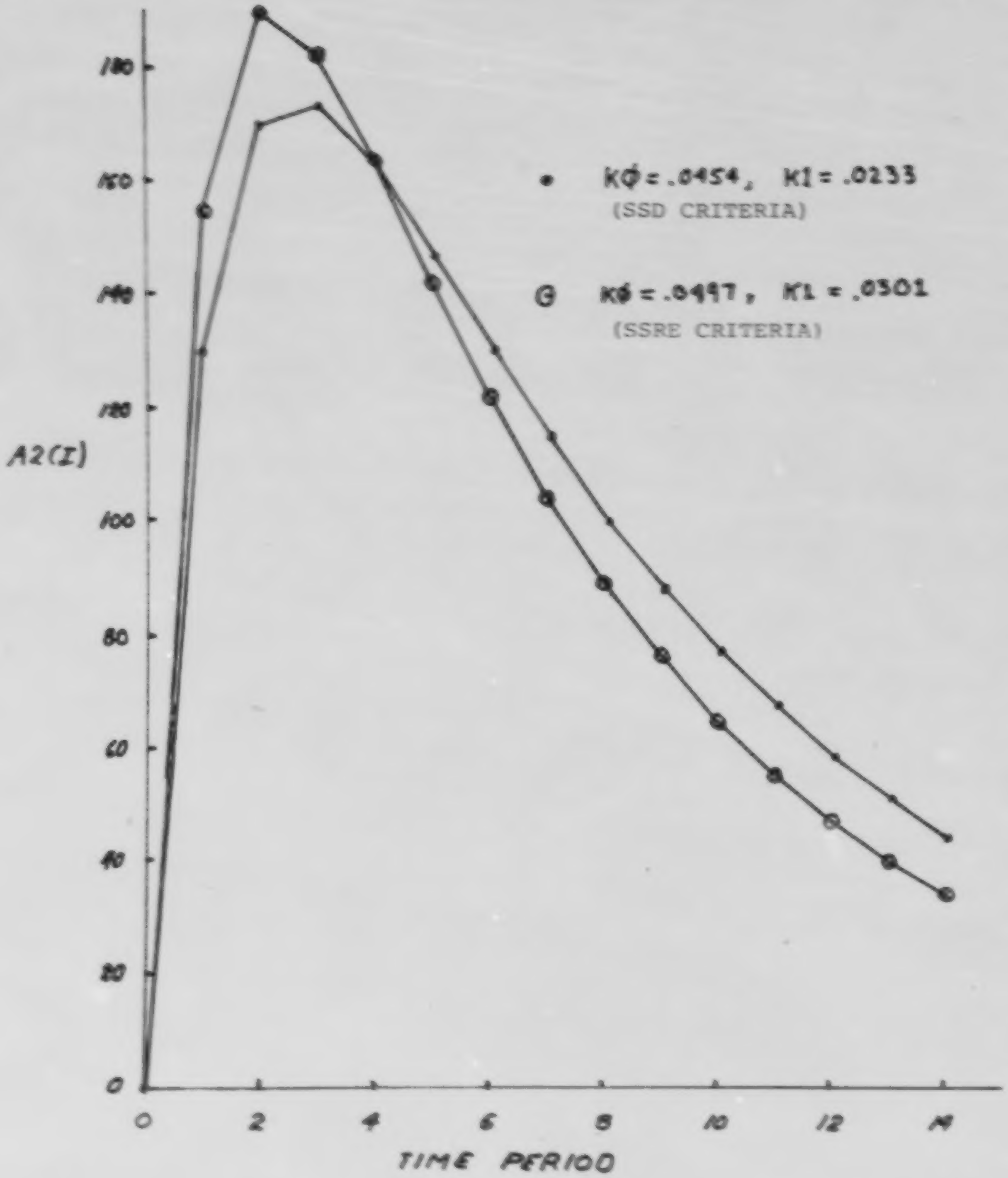


Fig. 9.25
Comparison of Concentrations in Extracellular Fluid

An Illustration of a Non-Linear Model

To demonstrate the flexibility and capability of the direct computer approach, your author altered the equation governing the time behavior of the insulin in the plasma compartment to include the assumption that the amount of insulin being transferred from the plasma to the urine in a time increment is equal to

A*A1(J)*A1(J)*D1.

In effect, this assumption states that the rate constant is proportional to the concentration. Mathematically speaking, one would say that we are assuming that the loss of insulin in a time period is proportional to the square of the concentration. Furthermore, by making this assumption, the fundamental character of the problem changes from a linear to a non-linear problem. It is nearly always the case that non-linear problems are intractable (unsolveable) with strictly formal mathematical techniques and that recourse must be made to computational methods. The quite ready solution of this far more difficult problem in a mathematical sense is yet another example of the flexibility of the computer based approach.

The <u>only</u> alteration required in the previous program is the replacement of line 710 with the statement:

```
710  LET A1(J+1) = A1(J)+(4*B*A2(J)-B*A1(J)-A*A1(J)*A1(J))*D1
```

This change was made in the program and the program was run using as initial guesses for the rate constants the values

K0 = 0.045   and   K1 = 0.023.

The program was unable to complete even one step in the search because of an overflow. Examination of the changed equation together with noting that the initial concentration for A1(0) was 3350, suggested that in the first time increment, the term

-A*A1(0)*A1(0)

was so large and negative that the value for A1(1) became negative. This resulted in the values of A1(2), A1(3) ... becoming even larger in magnitude, albeit negative; thus causing the overflow. This analysis suggested the initial selection of a much smaller value for K0 i.e. for A. The values

$$K0 = 0.0001 \quad \text{and} \quad K1 = 0.05$$

were then tried and after many search steps the value for K1 became negative. Because it is known that the rate constants must be positiv the calculation was programmed to stop should a rate constant become negative and hence the program did indeed halt.

Just to experiment, your author removed this restriction and permitted the program to run on. After several more steps the values of the rate constants became

$$K0 = 0.00002 \qquad \text{and} \qquad K1 = -0.00766.$$

The specific activity of the insulin in the second compartment, A2(5) was negative for all time increments. This too is a physiologically impossible result.

Your author included this discussion to give an example of the discretion which must be used in interpreting program results. The user must always keep in mind the constraints imposed by the "real world". He must not blindly accept the numbers that the program produces. The student should note that, as a result of the modification, the results produced were nonsensical This in turn is good evidence that the modification itself was wrong, i.e. the hypothesis that the amount of insulin leaving the plasma to the urin is proportional to the square of the concentration of the insulin in the plasma is false. It is thus necessary to alter the hypothesis and in this way the computer becomes a valuable tool for acquiring understanding of a phenomena because it enables a ready acceptance or rejection of hypotheses.

The problem of which hypotheses or models to accept or reject is particularly accute in the compartmental analysis of physiological systems. Atkins (1969) gives several examples of different models which are attempts to explain the same phenomena.

This concludes the portion of the chapter devoted to the compartmental analysis of physiological and/or biochemical phenomena. There are many other applications of compartmental analysis to these kinds of problems and they can be found in the journals and the texts listed in the references at the end of the chapter. The student who is familiar with electrical engineering should have noted the very close analogy between compartmental analysis and the lumped parameter method of analysis of electrical circuits.

The following sections will indicate some applications of compartmental analysis to ecological systems.

Food Chain Kinetics

In this section we consider the analysis of the movement of a small amount of some substance through a food chain. A food chain is a sequence of species and/or organisms through which food, energy or material is transferred as a result of one species eating another species and in turn being eaten by a third species, etc. Some examples of substances which are known to move through a food chain are pesticides, essential minerals and radionuclides. The determination of the concentrations of these substances in each species, as the substances move through the food chain, is a problem of great interest. As an example, the determination of the concentration of a radionuclide in the various components of a food chain provides a means of assessing the effect of the radioactivity on the individual components of the food chain.

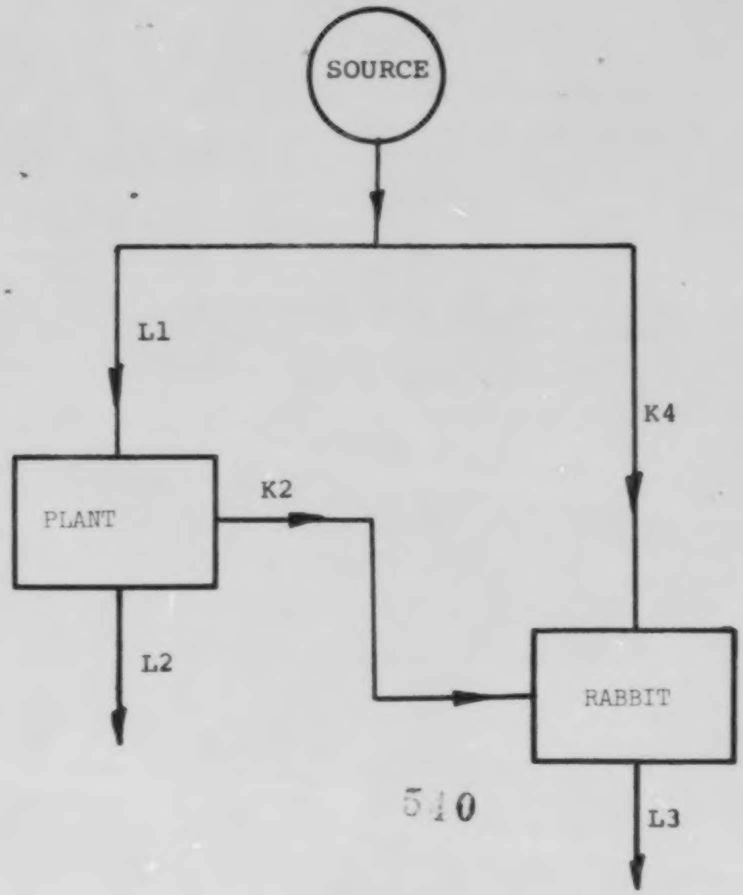
Because the laws governing the movement of small amounts of a substance through a food chain are very similar to the laws governing the behavior of the flow of tracers through physiological and biological systems, the method of presentation will closely parallel that used to discuss the previous examples. Emphasis will be placed upon the derivation of the equations and not upon the development of the computer programs. Since the derivation of the equations may occasionally require results from radiation physics or chemistry with which the student is not familiar, your author will simply state and use such results. The context of the derivation will usually indicate when such a procedure is being used. Despite the fact that the computer programs resulting from the quantitative description of food chain kinetics may appear quite straightforward, the obtainment of the necessary transfer coefficients, loss rates and other necessary data is frequently a difficult task. It is this fact which makes the obtainment of the fundamental equations a nontrivial matter.

A Nuclear Fallout Problem

Due to a nuclear test, nuclear fallout results in the deposition of radioactive material on an ecosystem. It is desired to describe the flow of the radionuclides through the system. We consider a very idealized form of this problem consisting of a single radionuclide, strontium-89, and a single ecosystem made up of a two-element food chain, plants, and rabbits. The strontium-89 is deposited on the plants which are in turn eaten by the rabbits and the problem is to determine the time variation of the concentration of the strontium-89 in both the plants and the rabbits. Since large concentrations of radioactivity are quickly lethal, it is of interest to restrict the discussion to very low levels of radioactive concentrations. For this reason it is usual to consider radioactive concentrations in units of microcuries (10^{-6} curies) or picocuries (10^{-12} curies). These magnitudes are comparable to those used in tracer studies.

This presentation will follow that given by L. Eberhardt, (1970) and the very difficult problem of acquiring accurate and sufficient empirical data will not be considered. Your author again cautions the student against building elaborate theoretical or computer models without proper consideration of experiemtnal and biological considerations.

Just as in the description of the behavior of a tracer in a physiological system, the description of the flow of a radionuclide in an ecosystem is accomplished by partitioning or subdividing the ecosystem into units or compartments, and then accounting for the amounts of radioactivity entering and leaving each compartment during a small increment of time. In this example, the ecosystem has been greatly simplified by imagining that there are only two compartments, the plants and the rabbits. A compartmental diagram is shown in figure 9.26.


SOURCE
L1
K4
PLANT
K2
L2
RABBIT
L3

Simple Two Compartment Nuclear Fallout Model

Fig. 9.26

The source of the radiation is the fallout of the strontium-89. The time variation of the resultant radiation intensity incident upon the ecosystem is accounted for by assuming that, the decrease in fallout radiation intensity in a small increment of time, D1, is directly proportional to the existing radiation intensity of the fallout and to the length of the time increment. In this, and the remaining problems, D1 is assumed to be very small compared to the total time of interest. Thus, if S(I) denotes the concentration of the incident source radiation present at the beginning of the I^{th} time increment, we have

$$S(I+1) = S(I) - L1*S(I)*D1 \tag{9.19}$$

where L1 is a constant of proportionality and is measured in units of inverse time. The equation governing the time variation of the concentration of the strontium-89 in the plants is derived from an accounting of both the amount of radioactivity absorbed by the plant in the small time increment as well as the amount of radioactivity leaving the plant during the same time increment. It is assumed that the amount of radioactivity entering the plant in the time period is proportional to the intensity of the radioactivity present from the fallout at the start of the time period, and to D1. The amount of radioactivity leaving the plant is assumed to be proportional to the intensity of the radioactivity present in the plant at the start of the period and also to D1. Hence, if P(I) denotes the intensity of the strontium-89 in the plant at the start of the I^{th} time increment, we can write

$$P(I+1) = P(I) + K2*S(I)*D1 - L2*P(I)*D1$$

or

$$P(I+1) = P(I) + (K2*S(I) - L2*P(I))*D1 \tag{9.20}$$

where K2 and L2 are constants of proportionality and L2 is called a loss rate. Both constants are measured in units of inverse time and the length of the time unit is the magnitude of D1.

The time variation in the intensity of the strontium-89 in the rabbit, R(I), is derived in a similar manner. In the time increment, D1, the concentration is increased by the amount of radiation incident upon the rabbit and the amount of radiation ingested by the rabbit from the plants. The amount of radioactivity incident upon the rabbit from the fallout is assumed to be proportional to the incident radiation intensity, to the exposed area A, of the rabbit, and to the length of the time increment. Furthermore, the amount of radiation ingested is assumed to be proportional to the amount of plant ingested during the time period, to the intensity of the radiation present in the plant at the time of ingestion and to D1. K3 and K4 denote the respective constants of proportionality and are expressed in units of inverse time. P1 will denote the amount of the plant eaten during the time increment and it will be assumed that the same amount is eaten each period. The student should note that the constants of proportionality relating the transfer of the radioactive Strontium-89 from the source to the rabbits and plants respectively have been purposely given the distinct labels L1 and K2. This has been done to emphasize the fact that the effective transfer rates of the radioactivity from the source to the plants and to the rabbits may be, and usually are, different. In animals, Strontium-89 usually concentrates in the bone marrow. The actual flow or passage of the radionuclide from the plant to the bone marrow is quite complex. The radioactive substance passes through the gastrointestinal tract and is taken up by the blood where it is then transferred to the bone marrow. The radioactivity leaves the bone marrow via the blood stream from which it is eventually excreted. Because a detailed representation of this process is not trivial, it is expedient to assume that the radioactive strontium-89 is transferred directly from the plant to the bone marrow and that the transfer is immediate. The process by which this transfer of radioactivity is effected is ignored and the bone is treated as a compartment. The assumption that the radioactivity is immediately transferred from the ingested plant to the rabbit could be removed by assuming a time lag for the radioactivity to pass from the ingested plant to the bone marrow of the rabbit. Finally, the loss in radiation from the rabbit during the increment of time is assumed to be proportional to D1 and to the radiation present in the rabbit at the start of the time period. The equation for the concentration of the radiation present in the rabbit is then given by

Equations 9.19, 9.20 and 9.21 represented the desired accounting of the entering and exiting Strontium-89 radiation. The amount of radiation incident from the fallout on the animal is so small that it may be neglected. In this event, equation (9.21) may now be written as

R(I+1) = R(I) + (K3*A*S(I) + K4*P1*P(I) - L3*R(I))*D1. (9.21)

R(I+1) = R(I) +(K5*P(I) - L3*R(I))*D1 (9.22)

where we have introduced the notation K5 = K4*P1.

The similarity between the derivation of these equations and the derivation of the equations governing the movement of tracers in physiological system should be quite evident. This very close analogy means that the programs used to solve each class of problems are also very similar. A program of this example is shown in figure 9.27 together with a listing of the results of a typical run. Statements numbered 250 and 510 enable the listing of data at integral units of time. The size of the dimension statements are such that the total number of time increments must be less than 105. Consequently, if for example it is desired to run the program for 100 integral units of time with an incremental time step of 0.2 units, the dimension statements will have to be enlarged to accommodate at least 500 entries. The loss and intake rates and the initial or starting values are taken from Eberhardt's paper.

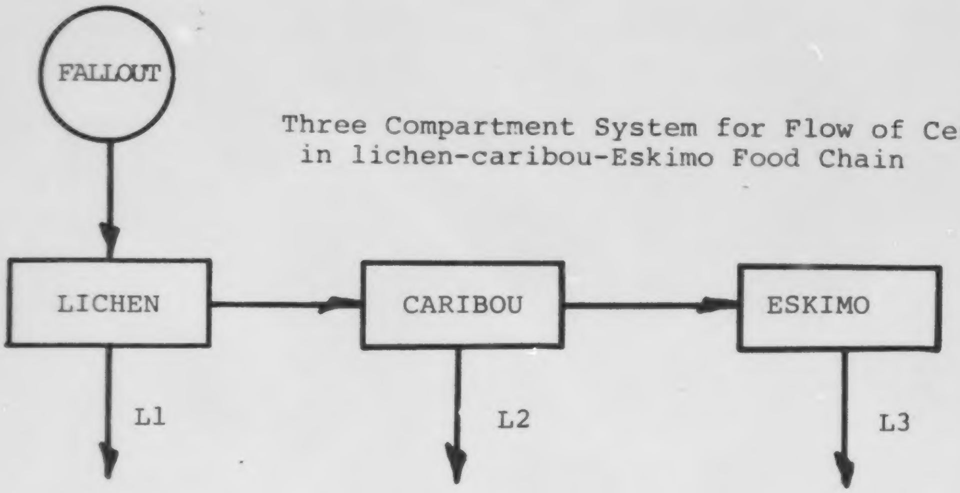
```
1 REM             JULY 30, 1975
5 REM   FLOW OF STRONTIUM-89 IN PLANT-RABBIT FOOD CHAIN
20 DIM S(105),P(105),R(105),T(105)
30 PRINT "TYPE  L1, THE FALLOUT LOSS RATE"
35 INPUT L1
40 PRINT "TYPE THE INTAKE RATE K2 AND THE LOSS RATE L2 OF THE PLANT"
45 INPUT K2,L2
50 PRINT "TYPE THE INTAKE RATE K5 AND THE LOSS RATE L3 OF THE RABBIT"
55 INPUT K5,L3
60 PRINT "TYPE  D1, THE LENGTH OF THE TIME INCREMENT"
65 INPUT D1
70 PRINT "TYPE  N, THE NUMBER OF TIME INCREMENTS"
75 INPUT N
200 LET S(0)=2750
210 LET P(0)=0
220 LET R(0)=0
230 REM  INSTR. 250  CALCULATES THE CORRECT PRINT INTERVAL
250 LET D2=INT(1/D1+1.00000E-06)
300 FOR I=0 TO N
320 LET S(I+1)=S(I)-L1*S(I)*D1
340 LET P(I+1)=P(I)+(K2*S(I)-L2*P(I))*D1
360 LET R(I+1)=R(I)+(K5*P(I)-L3*R(I))*D1
380 LET T(I+1)=T(I)+D1
400 NEXT I
495 PRINT
500 PRINT "I              T(I)          R(I)          P(I)         S(I)"
505 PRINT
510 FOR I=0 TO N STEP D2
520 PRINT I,T(I),R(I),P(I),S(I)
530 NEXT I
550 END
```

```
TYPE  L1, THE FALLOUT LOSS RATE
?.5
TYPE THE INTAKE RATE K2 AND THE LOSS RATE L2 OF THE PLANT
?1, .0346
TYPE THE INTAKE RATE K5 AND THE LOSS RATE L3 OF THE RABBIT
?.115, .0385
TYPE  D1, THE LENGTH OF THE TIME INCREMENT
?.1
TYPE  N, THE NUMBER OF TIME INCREMENTS
?100

I              T(I)           R(I)           P(I)           S(I)

 0             0              0              0              2750
 10            1              122.382        2169.73        1646.53
 20            2              432.455        3394.91        985.836
 30            3              837.721        4057.07        590.257
 40            4              1283.73        4384.57        353.409
 50            5              1738.76        4514.04        211.599
 60            6              2184.67        4527.22        126.692
 70            7              2611.39        4472.95        75.8552
 80            7.99999        3013.67        4380.42        45.4173
 90            9              3389.07        4267.03        27.193
 100           10             3736.84        4143.13        16.2815
```

Fig. 9.27

A Second Example in Food Chain Kinetics

This problem is adapted from the paper of Eberhardt and Hanson, 1969, which describes the time variation of the concentration of cesium-137 in a lichen-caribou-Eskimo food chain. A compartmental model of the system is shown in fig. 9.28.



Three Compartment System for Flow of Cesium-137 in lichen-caribou-Eskimo Food Chain

Fig. 9.28

The source of the cesium-137 radionuclide was the residue from nuclear weapons testing in the early 1960's.

The formulation of the problem proceeds in a manner entirely analogous to the procedure used in the preceding examples. Fick's principle is applied to each compartment to obtain the desired set of equations. The distinguishing features of this example are the detail required to mimic the time variation of the cesium-137 entering the food chain and the technique used to obtain the transfer coefficients.

Because records were not available for the accumulation of cesium-137, Eberhardt and Hanson used the monthly accumulation records of Strontium-90 fallout as recorded in Fairbanks, Alaska and shown in table 9.2. It was assumed that the ratio of the intensity of the cesium-137 to the intensity of the strontium-90 remained constant throughout the time period during which the data was collected and hence the rate of deposition of the cesium-137 would be identical to the rate of deposition of the strontium-90. In this way the time variation of the cesium-137 entering the food chain could be estimated.

Estimates of strontium-90 deposition
at Fairbanks, Alaska (mCi/km^2)

(from Eberhardt and Hanson, 1969)

Year	Jan.	Feb.	Mar.	Apr.	May	June	July	Aug.	Sept.	Oct.	Nov.	Dec.	Total
1962	0.08	0.22	0.18	0.08	0.22	0.99	0.76	0.66	0.13	0.07	0.25	0.04	3.68
1963	0.15	0.35	0.60	0.40	0.30	2.43	2.54	1.62	0.19	0.17	0.02	0.07	8.84
1964	0.06	0.11	0.07	0.22	0.44	1.37	0.01	0.61	0.14	0.12	0.09	0.07	3.31
1965	0.04	0.02	0.14	0.11	0.07	0.55	0.25	0.29	0.14	0.28	0.06	0.10	2.05
1966	0.03	--	0.03	--	0.13	0.12	0.09	0.06	--	0.02	0.04	0.01	0.53

--: zero to trace

Table 9.2

The choice of the units with which to express the intensity of the radionuclide is set by the ease with which the necessary parameters and data may be obtained and by convenience. Thus, the intensity of the cesium-137 radiation in the lichen compartment was expressed in nano curies per square meter, i.e. nCi/m^2. This choice was partially dictated by the fact that it was experimentally possible to estimate the conversion coefficient relating this unit to the radiation intensity per gram dry weight of caribou muscle. The latter unit was selected because it permitted the determination of a conversion factor from nCi/gm dry weight of caribou muscle to units of total radiation, or body burden, in the Eskimo. This individual choice of units was also very convenient because it avoided the necessity for estimating the amount of lichen eaten each day by the caribou.

In deriving the equation for the time variation of the cesium-137 in the lichen, it is necessary to obtain an expression for the amount of radiation entering the lichen in a small time period as well as an expression for the amount of radiation leaving the lichen during the same period of time. The latter expression is easily obtained by assuming that the decrease in cesium-137 intensity in the lichen is proportional to the magnitude of the intensity in the lichens. The obtaining of an expression for the amount of radiation entering the lichen requires a knowledge of the time variation of the cesium-137 fallout intensity. This time variation was not recorded; however, the strontium-90 fallout intensity was recorded at regular time intervals over a period of several years. By assuming that the ratio of the cesium-137 intensity to the strontium-90 intensity remained constant over the entire period of recording, it follows that the time variation of the intensity of the cesium-137 is the same as that of the strontium-90. Hence, an equation describing the intensity of the strontium-90 over the time period of interest will be developed and this equation will be used to give the time variation of the cesium-137. The equation will be derived by assuming a model for the time variation and then determining the parameters in the model so that the equation 'best' approximates the recorded data. The development is given in the next few paragraphs.

An analysis of the strontium-90 fallout data as depicted in table 9.2 reveals a monthly variation in the fallout intensity. Nevertheless, the data "appears" to indicate that the magnitude of the intensity is decreasing with time. The monthly variations can be somewhat removed by calculating the yearly accumulations. This is a very crude form of data smoothing. If the year is assumed to extend from June 1 to June 1, the yearly accumulations are given in table 9.3:

YEAR	ACCUMULATION
1963-1964	7.94
1964-1965	2.75
1965-1966	1.86
1966-1967	0.45

The last figure is the result of a very cavalier extrapolation from the monthly data. These tabular values clearly indicate the decrease of the intensity with time. Moreover, the decreasing set of values appear to be exponentially decreasing. Recalling that radioactivity decays in an exponential manner, further suggests that a reasonable approximation to the time variation of the fallout intensity may be obtained by assuming that the decrease in the intensity is proportional to the existing intensity. Thus, we write

$$F(I+1) = F(I) - A*F(I)*D1 \qquad (9.23)$$

where F(I) is the daily radiation intensity, D1 is chosen to be one day and A is a decay rate to be determined by parameter fitting.

F(I) will be expressed in units of nCi/m^2. The parameter A will be determined by using equation 9.23 as a basis for calculating the daily fallout intensities which in turn will enable the calculation of the total fallout each year. The total fallout each year will be compared to the empirical values appearing in tabel 9.3 using the sum of the squares of the deviations closeness criteria. The value of A, which results in the smallest sum of the squares of the deviations, will be the desired decay rate. The magnitude of the daily fallout intensity is the decrease in cesium-137 intensity from one day to the next. This decrease is equal to A*F(I)*D1 where F(I) is calculated according to equation 9.23. The sum of such terms for I=0 to 364 is the accumulated fallout intensity in a single year.

A program which performs these calculations, together with the results of a typical run, is shown in figure 9.29. Equation 9.23 is represented by line 220 of the program. The accumulation of fallout in a single year is accomplished by inserting the equation

$$Y = Y+A*F(I)*D1 \qquad (9.24)$$

in the same loop in which the daily variation in fallout intensity is calculated. The loop consists of instructions 200 to 260 and line 210 corresponds to equation 9.24. The student should recognize that the form of equation 9.24 is the standard form for accumulating or summing a set of quantities. In this instance, the quantities are the terms A*F(I)*D1 for I = 0,1,2, ,364. Line 280 stores the yearly accumulated fallout for later printing and line 300 calculates the deviations of the empirical and calculated yearly fallout intensities. Instruction 320 enables the summing of the squares of the deviations and line 340 insures that the previous years accumulated fallout radiation will not be carried over to the next years total.

```
RBC24

10 REM    DET. OF FALLOUT DECAY CONSTANT IN ALASKA FOOD CHAIN
11 REM
20 PRINT "TYPE THE INITIAL INTENSITY, F"
25 INPUT F
26 PRINT
30 PRINT "TYPE  A,  THE ESTIMATED VALUE OF THE DECAY RATE"
35 INPUT A
36 PRINT
37 REM
38 REM              LINES 40-55 INPUT THE YEARLY INTENSITY
39 REM
40 LET E(1)=7.94
45 LET E(2)=2.75
50 LET E(3)=1.86
55 LET E(4)=.45
139 REM
140 REM      THE OUTER J LOOP CALCULATES THE MEASURE OF CLOSENESS
141 REM
150 REM     LINES 200-260 CALC. RAD. INT. IN ONE YR. USING D1=1 DAY
151 REM
160 FOR J=1 TO 4
190 LET D1=1
200 FOR I=0 TO 364
210 LET Y=Y+A*F*D1
220 LET F=F-A*F*D1
260 NEXT I
280 LET C(J)=Y
300 LET D2=Y-E(J)
320 LET S=S+D2*D2
340 LET Y=0
360 NEXT J
400 PRINT "THE SUM OF THE SQUARES OF THE DEVIATIONS IS";S
405 PRINT
410 PRINT " J                    C(J)           D(J)"
415 PRINT
420 FOR J=1 TO 4
430 PRINT J,C(J),E(J)
440 NEXT J
500 END

READY

       TYPE THE INITIAL INTENSITY, F
       ?13

       TYPE  A,  THE ESTIMATED VALUE OF THE DECAY RATE
       ?.0025

       THE SUM OF THE SQUARES OF THE DEVIATIONS IS .534489

        J               C(J)           D(J)


        1               7.78621        7.94
        2               3.12274        2.75
        3               1.25241        1.86
        4                .502292        .45
```

Figure 9.29

It is necessary to provide a starting source intensity, F(0). Because this value must also be determined, the problem is actually a two-parameter determination problem. The program shown in figure 9.29 enables the guessing of the parameter values. Your author did not use an automated search routine because he was only interested in illustrating a procedure, not in the obtainment of refined numbers such as would have been produced by an automated search routine. It also would have been of interest to use a different closeness criteria such as the sum of the squares of the relative errors. A few trial runs resulted in a value of 13.0 nCi/m^2 for the initial source intensity and a value of 0.0025 $(DAY)^{-1}$ for the parameter A. The variation of the source intensity may then be expressed as

$$F(I+1) = F(I) - 0.0025*F(I) \tag{9.25}$$

where $F(0) = 13.0\ nCi/m^2$. This is the desired equation for the strontium-90 fallout intensity and completes our discussion on how to obtain it. By recalling that the strontium-90 fallout intensity is 1.7 times the cesium-137 fallout intensity, it is possible to use equation 9.25 to describe the time variation of the cesium-137 fallout intensity. This may be done by using a proper starting value, F(0), which corresponds to the initial cesium-137 radiation intensity.

The radiation transferred from the lichen to the caribou depends upon the amount of lichen eaten by the caribou which in turn depends upon the average grazing area of the caribou. Thus, a knowledge of the intensity of the fallout radiation upon a unit area of the lichen, in conjunction with the magnitude of the grazing area of the caribou, will enable an estimate of the consequent radiation transferred from the lichen to the caribou. The unit of area was chosen to be one square meter. Now only a certain fraction, K, of the fallout radiation coming to rest upon the lichen is actually absorbed into the lichen. This means that not all of the fallout intensity, F(I), will be transferred to the lichen. However, because not all of the incident fallout is transferred to the lichen, the actual amount of strontium-90 absorbed into the lichen is K*F(I) where K is a proper fraction. One method for estimating the transfer coefficient, K, is based upon an analysis of the fundamental equation governing the time behavior of the radiation intensity in the lichen.

To derive this equation, we begin by noting that the amount of radiation entering the lichen in a time period is

K*F(I)*Dl.

If L(I) denotes the intensity of radiation in the lichen during the Ith time period, by assuming that the loss in radiation is proportioned to the radiation, the term

-Ll*L(I)*Dl

is seen to represent the amount of radiation lost by the lichen during the same period. Ll is a constant and is assumed to be 0.0002 and is obtained from experiment. Ll represents the fraction of the existent lichen intensity lost during the time period. The loss rate from the lichen is less than the natural decay rate of cesium-137 because there exist many mechanisms whereby the lichen can lose radiation other than by just the normal radioactive decay of the cesium-137. Some of these mechanisms are: the washing away of the radiation on the surface of the lichen by rainfall, the carrying away of the radiation by evaporation, the loss of the radiation to the soil, etc. For these reasons there has been a range of estimates of the loss rate varying from 0.0001 to 0.0004. The lichen radiation intensity equation is obtained by an application of Fick's principle to yield

$$L(I+1) = L(I)+ (K*F(I)-L1*L(I))*D1. \tag{9.26}$$

With the aid of this equation we can now obtain a value for K. We begin by recalling the work with the prey-predator models. In that work, a maximum or a minimum of the population was noted to occur when there was no change in the population variable from one time period to the next. Thus, for the particular time period corresponding to a maximum value of the cesium-137 radiation intensity in the lichen, it must be the case that there is no change in the intensity. Hence

$$L(I+1) = L(I)$$

or

$$K*F(I) - 0.0002*L(I) = 0 \tag{9.27}$$

for that particular time period, I. (These two equations do not hold for all time periods). From observation, it was noted that the maximum concentration in the lichen was 30 pCi/gm. Since there are about 1600 gr/m^2 of dry weight of lichen, the maximum concentration of radiation in the lichen is 48,000 pCi/m^2 or 48 nCi/m^2. Now, an analysis of the data reveals that the maximum radiation concentration in the lichen occurred about three years after the start of the collection of fallout data. The approximate value of the maximum intensity of the fallout radiation was 0.84 nCi/m^2. This value was obtained by using equation 9.25 as the basis of a program which was run for three years or 1095 time increments with an initial fallout intensity of 13.0 nCi/m^2.

With the aid of these results, equation 9.26 may now be written as

$$K*0.84 - 0.0002*48 = 0,$$

which enables the value of the transfer coefficient to be calculated as

$$K = 0.011$$

It was assumed that the initial concentration of cesium-137 in the lichen was 15 pCi/gm dry weight or 24 nCi/m^2. This value, together with the value of K permits the use of equation 9.26 in describing the time variation of the radiation intensity in the lichen.

Because it is assumed that most of the radiation body burden in the Eskimo is due to the eating of caribou muscle by the Eskimo, we require an equation governing the transfer of cesium-137 through the caribou muscle. The unit of radiation intensity in the muscle was chosen to be nCi/gm dry weight of caribou muscle. This unit was chosen because it enables an easier conversion to the entering radiation for the Eskimo radiation intensity equation. It is known that the process by which the radionuclide enters the caribou muscle from the lichen is complicated. There is also a time lag because the radiation does not immediately concentrate in the caribou muscle after the caribou has eaten the lichen. Both of these factors are ignored in our model. The former factor is ignored because we are interested in the overall behavior of the entire food chain and the latter effect is ignored because the time of interest scale is very large compared to the elapsed time from the time the lichen enters the caribou to the time the radiation in the lichen has concentrated in the caribou muscle. The basic equation governing the time variation of the intensity of the radionuclide in the caribou muscle is very similar to the previous equation for the time variation of the radionuclide intensity in the lichen. The equation is

$$C(I+1) = C(I)+(C1*L(I) - L2*C(I))*D1 \tag{9.28}$$

where C1 is a conversion coefficient relating the effective transfer of the radiation intensity in the lichen to the intensity in the caribou muscle and L2 is an effective loss rate. Just as with the lichen, there are many mechanisms by which a living organism can lose radiation other than by the process of pure radioactive decay of the radionuclide. Consequently, the student will note that in this example, as in other food chain kinetic problems, the loss rates are effective loss rates and not pure radioactive decay rates.

To obtain a value for L2, we recall from a knowledge of radiation chemistry and experiment, that in 30 days the radiation intensity would be reduced by one-half. In the literature such a loss time is referred to as a half-life or a half-time. Now, by using the equation of radiation loss,

$$R(I+1) = R(I) - L2*R(I)*D1,$$

setting R(0) = 1, assuming a value for L2 and then running the program for the number of time increments necessary to yield a value of R(I) equal to one-half of R(0), the effective half-time can be determined. The half-time is the value of I expressed in units of D1. At first sight this procedure does not seem to be of much help in the determination of L2 since a value of L2 is required to use the program. However, by recalling some of the problems at the end of the first chapter dealing with exponential growth, it is possible to use such a program to determine the value of L2 which results in a specified half-time. This can be done by assuming an initial value for L2 and then, using the program to perform the calculations, an accurate value for L2 can be obtained by trial and error. With D1 set equal to one day, it is found that the loss rate corresponding to 30 days is approximately 0.02 $(\text{day})^{-1}$. The student should write a program and attempt to verify this. As used in the above equation, L2 is sometimes called a decay parameter.

Upon digestion of the lichen by the caribou not all of the radiation in the lichen is transferred to the caribou. Thus, C1 is an effective transfer coefficient relating the effectiveness of the transfer of radiation from the lichen to the caribou. This conversion coefficient depends upon the amount of lichen eaten each day, as well as on the net amount of the radiation in the digested lichen ultimately residing in the caribou muscle. A value of 0.1 was thought to be a representative value. One possible method for obtaining a conversion value is to permit the caribou to graze on lichen having a known and constant radiation intensity. The length of the grazing period is assumed to be such that the radiation intensity in the caribou muscle has reached an equilibrium level. The intensity levels in both the caribou muscle and the lichen are then noted and an effective loss rate is assumed for the radiation in the caribou muscle. This loss rate was set at 0.02 and with a measured equilibrium intensity of 200 pCi/gm dry weight in the caribou muscle it is evident that the radiation excretion was 0.02*200 pCi/gm or 4 pCi/gm dry weight of caribou muscle. The measured equilibrium intensity of the lichen was 40 pCi/gm dry weight and, therefore, the intensity of the caribou excretion loss was about 0.1 of the radiation intensity in the lichen. Because the flow of radiation from the lichen to the caribou was assumed to be in equilibrium, that is in a steady state, the gain in radiation intensity from eating the lichen must equal the loss in radiation intensity due to excretion loss. For this reason, the value of 0.1 was chosen for a conversion value to convert from existing lichen radiation to entering caribou radiation.

With the aid of the above results, equation 9.28 may be written as

$$C(I+1) = C(I)+(0.1*L(I) - 0.02*C(I))*D1. \tag{9.29}$$

The initial concentration of cesium-137 in the caribou muscle was set at 80 pCi/gm dry weight.

The Eskimo radiation is

$$E(I+1) = E(I)+(E1*C(I) - L3*E(I))*D1 . \tag{9.30}$$

This equation is derived by assuming that the gain in body burden radiation is proportional to the intensity of radiation in the caribou muscle and the loss in radiation body burden is proportional to the magnitude of the body burden radiation which is expressed in units of nCi. The effective loss rate, L3, was chosen equivalent to a 70 day half-life and hence $L3 = 0.01(\text{day})^{-1}$. This value for L3 was obtained in a manner similar to the method for obtaining L2. The estimate of the conversion factor from caribou muscle radiation intensity to Eskimo body burden was based upon an assumption of constant meat intake for each individual and required a comparison of computed caribou muscle radiation intensity with observed Eskimo body burdens. Because of the complications of obtaining complete data, as well as the difficulty of obtaining a reliable estimate of the conversion parameter, we will simply postulate that a reasonable value for the Eskimo body burden conversion factor is 80. For complete details on the obtaining of such an estimate, see the paper by Eberhardt and Hanson previously referred to, as well as the paper by Hanson and Eberhardt, March 1969. This latter paper also provides evidence that a reasonable value for the initial Eskimo body burden is 500nCi.

Equations 9.25, 9.26, 9.29, and 9.30, together with the prescribed initial conditions, permit the construction of a computer program which describes the time variation of the cesium-137 radiation intensity in each compartment.

In their model, Eberhardt and Hanson included the time variation of the amount of lichen eaten by the caribou. This can be done by multiplying the conversion coefficient by a factor which changes in time. The authors' state, "On the premise that some lichen is present in the diet at all times, minimum content was set at 10% and increased very gradually for about five months, reaching about 30% in the fifth month, and then rising rapidly (in a few weeks) to 100% and remaining at that level for six months." The variation is repeated each year. The authors also accounted for a time variation in the Eskimo diet of caribou meat. We again quote from the Eberhardt and Hanson paper. "Thus a 'spring kill' is assumed to take place on June 1, and intake concentrations held at that level for 2 months, dropped to one-half the same level for 15 days (to correspond to a reduced supply of meat), then to zero for a further two months, whereupon a new sampling is made (to correspond to a mid-October caribou harvest) and utilized as intake until the following June 1, when the cycle is repeated." This variation can also be introduced into the model with the aid of a multiplicative factor, which is to vary in time and magnitude according to the previous statements.

In this example the student should note that a good portion of the development was devoted to obtaining the conversion coefficients. These coefficients are measures of the effectiveness of the transfer of radiation from one compartment to another and for this reason could also be called transfer coefficients. The methods for obtaining the coefficients are very useful and so we recapitulate them. The first method relies on the use of the steady state or equilibrium condition. In this condition it is postulated that the gain in radiation intensity in a time increment is equal to the loss in radiation intensity during the same time increment, that is, the amount flowing in is equal to the amount flowing out. Now, in our models of the time variation of the radiation intensity, the expression for the amount flowing in, as well as the expression for the amount flowing out, are all of the same form for each compartment.

The amount flowing in, in a unit of time, is given by an expression of the form C*C(I) and the amount flowing out in the same unit of time is given by a similar expression, L*L(I). By setting up an experiment in which it is assumed that the flow between compartments is in the steady state it is frequently possible to measure C(I) and L(I). Since the loss coefficient, L, is also known or assumed, the equality of the two expressions provides a method of obtaining C, the transfer coefficient. Thus

$$C = L*L(I)/C(I). \tag{9.31}$$

The second method begins with the observation that the expression C*C(I) - L*L is the change in radiation intensity in a compartment in a time period. Suppose that the experimental determination of the values of C(I) and L(I) at the time of maximum or minimum radiation intensity of the compartment can be effected, then th use of the equation

$$C*C(I) - L*L(I) = 0$$

enables the evaluation of the transfer coefficient. This is due to the fact that, at a time of maximum or minimum radiation intensity in the compartment, the change in the intensity is zero.

Finally, by assuming different values for the transfer coefficient, and comparing the calculated compartmental radiation levels with those measured experimentally, it is also possible to determine the transfer coefficient. There are thus three different methods for the determination of these coefficients.

The program for this model will not be developed since the preceding discussion should be sufficiently detailed to permit the student to write his own program. We leave this as a problem.

The Movement of a Pesticide in a Food Chain

Because of the abundant use of pesticides in the raising of food crops, and the associated possible long term effects of pesticides, the description of the movement of a pesticide through a food chain is of current interest. This example, which is adapted from O'Neil and Burke, 1971, describes the development of a very simple compartmental model for the description of the movement of DDT and DDE through a human food chain. Your author chose this example because it serves to illustrate the required balance between model sophistication and practical considerations. The length of time in which the authors were permitted to develop the model was very short and also the paucity of the data permitted the determination of only a few model parameters. These constraints necessitated the development of a simple model as well as the exercise of caution in the interpretation of the model results.

The data shown in table 9.4 lists the yearly pesticide concentrations in the food and in the human together with the intensity of the pesticide usage. The data is taken from the above-mentioned work of O'Neill and Burke. The intensity of the pesticide in the human was measured in the human adipose fat tissue. An examination of the data suggests a decline in

Table 9.4. Data utilized for modeling DDT and DDE movement through the human food chain.[a]

Year	DDT Usage (10^6 lb)	DDT + DDE in Market Place Diet (mg/kg)	DDT + DDE in Human Adipose Tissue (ppm)[d]
1965	53	0.031	
1966	46	0.040	
1967	40	0.026	4.65[e]
1968	33	0.019	5.61
1969		0.016	5.22
1970		0.015	5.27

---

[a] From O-Neill and Burke, 1971

the usage of the pesticide as well as a decline in the concentration of the pesticide in the food in the market place. The pesticide concentration in humans appears to oscillate; however, as the authors point out in their paper, there is some justification for questioning the accuracy of the data. It is evident that a decrease of the 1970 intensity data by only two-tenths of one per cent would result in the appearance of a continual decline in the concentration of the pesticide in the human from the peak year of 1968. This overall change of the character of the data due to such a small change in the data, coupled with the fact that the reliability of the data was difficult to assess, further mitigates against the construction of an elaborate model.

The determination of the compartmentalization, as well as the number of compartments, is one of the first steps in the construction of the model. This initial determination is an art and depends to a great extent on the intuition and insight of the investigator. In keeping with the policy stated by Atkins, 1969, concerning the number of compartments to be used in developing tracer models, it is desirable to use a minimum number of compartments consonant with the obtaining of "reasonable" agreement with the data. Thus, in this problem, a two-compartment model was initially assumed. The compartments consisted of the food supply and the human adipose tissue. Since it was not possible with this two-compartmental model to obtain reasonable agreement with the qualitative behavior of the data, a three compartment model was hypothesized. The authors divided the food supply into two compartments, each of which were distinguished by their rate of absorption of the pesticide. The first compartment was assumed to account for the short term absorption of the pesticide, e.g. the effect of direct spraying of the pesticide upon the food supply. The second compartment was postulated to account for the long term effect of the pesticide through recycling from the soil of the pesticide sprayed upon the soil.

By using these three compartments, O'Neill and Burke were able to develop a model which mimiced the qualitative behavior of the data. The three compartment model is shown in figure 9.30 on the next page.

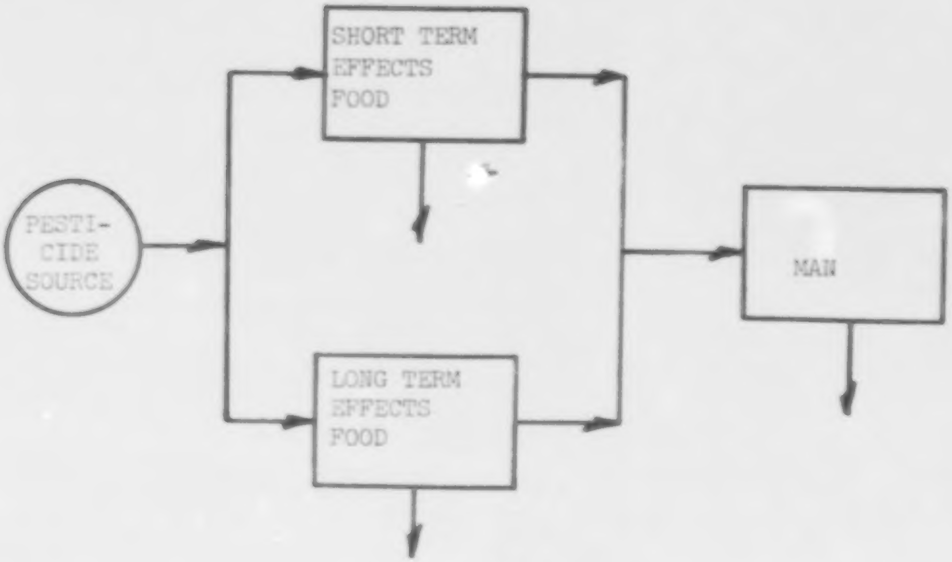


Fig. 9.30

The derivation of the governing equations follows from an application of Fick's principle to each compartment together with the assumption that in a small time increment the amount of pesticide gained by a compartment is proportional to the concentration of the pesticide of the inputting or feeder compartment and the amount of pesticide lost in the same time increment is proportional to the concentration of the pesticide in the compartment. The constant of proportionality relating the amount or change in concentration of the pesticide from a compartment to its successor compartment is best thought of as a transfer coefficient or a conversion factor. This is because not all of the pesticide leaving a compartment 'effectively' enters the successor compartment. The situation is quite analogous to the transfer of radionuclides from one compartment to the next. The constant of proportionality relating the existing concentration of pesticide to the amount leaving the compartment in an increment of time is also an effective constant because it accounts for all possible ways the pesti may leave the compartment other than by the sole transfer of the pesticide to another compartment.

Let S(I) denote the quantity of the pesticide expressed in millions of pounds, F1(I) and F2(I) denote the concentration of the pesticide in the respective food compartments expressed in milligrams of pesticide per kilogram of weight of food, i.e. mg/kg and finally let H(I) denote the magnitude of the pesticide concentration in the human adipose tissue measured in parts per million, ppm. In terms of this notation, the governing equations may be written as:

$$F1(I+1) = F1(I) + (K1*S(I) - L1*F1(I))*D1 \qquad (9.32)$$

$$F2(I+1) = F2(I) + (K2*S(I) - L2*F2(I))*D1 \qquad (9.33)$$

and

$$H(I+1) = H(I) + (K3*(F1(I) + F2(I)) - L3*H(I))*D1 \qquad (9.34)$$

and D1 is chosen to be one day.

An expression for S(I), the daily usage may be obtained by curve fitting and/or interpolation. Curve fitting is a process whereby one function, or a set of data, is represented by another function. In this section, we introduce a simple type of curve fitting which is equivalent to approximating the data by a straight line. The time variation of the DDT usage appears to be such that the usage decreases by a constant amount each year, that is, the quantity (S(I+1) - S(I)) is constant from year to year. This suggests that the time variation of the intensity can be approximated by the expression

$$S(I+1) = S(I)+S1*D1 \qquad (9.35)$$

where S1 is a constant to be determined so that the values of S(I) are close to the empirical values which we will denote by E(I). The closeness criteria will be chosen to be the least squares criteria and thus we will attempt to determine S1 so that the sum of the terms

$$(S(I)-E(I))^2$$

for I = 0, 365, 730, and 1095 is a minimum. These numbers correspond to 0, 1, 2 and 3 years respectively because D1 is chosen to be one day. This also implies that S1 is expressed in units of the number of millions of lbs. per day. It should be evident to the student that this method of obtaining an expression for approximating the source usage is very analogous to the method employed in the previous problem to obtain an approximating expression for the empirically determined fallout intensities. The program to accomplish this is very similar to that given in figure 9.29 and hence will not be reproduced here. The initial value, S(0), can be made to coincide with the recorded usage in the year 1965, or the initial value can be left as a free parameter to be determined. Your author obtained the value of S1 = -0.018 under the constraint of initially setting S(0) = 53.0 and he obtained the value of S1 = -0.019 and S(0) = 54 if S(0) was also determined as a free parameter. The value of 365*(-0.019) is approximately -7.0 and this implies that the value of -0.019 for S1 corresponds to a yearly decrease of seven million pounds of usage of the pesticide which rate of usage is in accordance with the tabular data.

The constancy of the expression (S(I+1) - S(I)) for all increments of time means that the graph of S(I) versus I appears as a straight line. Hence, the term linear or straight line approximation for such a representation of the data. In the Alaska food chain example, the strontium-90 fallout intensity was approximated by equation 9.23. If this equation is written in the form

$$(F(I+1) - F(I))/F(I) = -A*D1,$$

is seen that exponential approximation corresponds to assuming that the relative change in F is constant. In contrast, linear approximation assumes that the magnitude of the change in any time increment is constant. These are the two assumptions most frequently used in describing data. A surprisingly large and varied number of phenomena are quite well approximated by one or the other of these approximations.

The determination of the six parameters, K1, K2, K3 and L1, L2, L3 is accomplished by a comparison of model and empirical results. This method of determination is in contrast to the method used to obtain the parameters in the two previous examples. In those examples other information was available which permitted an easier and more direct determination.

The fact that a two parameter model could not be constructed which would suitabl match the empirical data means that it was not possible to find values for the parameters for the two parameter model such that the model results agreed with the empirical data. The student should realize the limitations of the implication of this statement. All of the models were constructed using Fick's principle and the assumption that the fluxes were proportional to the concentrations. Hence, if for example, the fluxes had been assumed to be proportional to the square of the concentrations it might have been possible to have obtained agreement with the use of only a two-parameter model. Other two parameter models could have been constructed utilizing still other assumed relations between the fluxes and the concentration in the compartments. Thus, the statement "that reasonable agreement could not be obtained for a two parameter model" actually means that, for the special case of a linear compartment model, that is, a model for which the fluxes were assumed to be proportional to the first power of the concentration, it was not possible to obtain reasonable agreement between the model results and the data. The assumption that the fluxes are proportional to the first power of the concentration corresponds to the assumption of first order kinetics as used in describing chemical reactions.

The transfer coefficients and the loss rates must be obtained from a multidimensional search routine. Because there are six values to be determined, it is not feasible to use such a routine unless the investigator has a "reasonable" estimate of starting values for the parameters. To get such values, your author first wrote a program for calculating the time variation of the pesticide concentrations in the three compartments assuming that the values of the parameters were specified. Provision was made to specify these values as input and the sum of squares closeness criteria enabled the determination of the measure of closeness for any given set of values for the parameters. Since the parameter values were specified as input, it was possible "to play with alterations of the parameters to get a feel" for some reasonable values to use as initial values for the multidimensional search program. The program for calculating the pesticide concentrations is given in figure 9.31, and the results of a typical run are displayed in figure 9.32. The student will note that lines 290, 310 and 330 correspond to equations 9.31, 9.32 and 9.34 respectively. In the text, subscripted variables are used for ease of reading by the student; however, they are not necessary in the program. The index J counts the years and the index I counts

```
350 LET S1(J)=S
370 LET N1(J)=F1
390 LET N2(J)=F2
410 LET H1(J)=H
500 NEXT J
530 PRINT " J   S1(J)     N1(J)            N2(J)          N1(J)+N2(J)       H1
(J)"
532 PRINT
550 FOR J=0 TO 4
570 PRINT J;S1(J),N1(J),N2(J),N1(J)+N2(J),H1(J)
590 NEXT J
699 REM
700 REM    LINES 705-720 ENTER FOOD PESTICIDE CONCENTRATIONS
701 REM
705 DATA 40,26,19,16,15
710 FOR I=0 TO 4
715 READ E3(I)
720 NEXT I
724 REM
725 REM      LINES 730-745 ENTER HUMAN PESTICIDE CONCENTRATIONS
726 REM
730 FOR I=0 TO 4
735 DATA 2,4.65,5.61,5.22,5.27
740 READ E4(I)
745 NEXT I
799 REM
800 REM     LINES 800-840 CALC. M1, THE MEASURE OF CLOSENESS
801 REM
810 LET M1=0
820 FOR W=0 TO 4
830 LET M1=M1+(N1(W)+N2(W)-E3(W))^2+(H1(W)-E4(W))^2
840 NEXT W
845 PRINT
850 PRINT
855 PRINT "THE VALUE OF THE CLOSENESS CRITERIA IS";M1
1000 END
```

Figure 9.31 (Cont.)

```
RBC35

5 REM  PESTICIDE FOOD CHAIN PROBLEM
6 REM
7 REM
20 PRINT "TYPE THE INFLUX TRANSFER COEFFICIENTS  K1,  K2,  K3"
25 INPUT K1,K2,K3
30 PRINT
35 PRINT "TYPE THE LOSS RATES  L1,  L2,  L3"
40 INPUT L1,L2,L3
42 PRINT
50 PRINT "TYPE THE INIT, VAL. FOR FOOD PEST. CONC., F1, F2"
55 INPUT F1,F2
56 PRINT
60 PRINT "TYPE  H, THE INIT. CONC. OF PEST. IN THE HUMAN"
63 INPUT H
64 PRINT
65 PRINT
66 PRINT
67 REM
68 REM     LINES 70-80 STORE INITIAL VALUES FOR LATER USE
69 REM
70 LET N1(0)=F1
75 LET N2(J)=F2
80 LET H1(0)=H
84 REM
85 REM  INSTRS. 95-110  INITIALIZE VARIABLES
86 REM
95 LET D1=1
100 LET S=53.0001
110 LET S1(0)=S
144 REM
145 REM INSTRS. 150-500 ARE MAIN PART OF CALCULATION
146 REM
150 FOR J=1 TO 4
239 REM
240 REM  INSTRS 250-340 DAILY UPADATE THE VARS. FOR A YEAR'S TIME
241 REM
250 FOR I=0 TO 364
270 LET S=S-.0185*D1
290 LET F1=F1+(K1*S-L1*F1)*D1
310 LET F2=F2+(K2*S-L2*F2)*D1
330 LET H=H+(K3*(F1+F2)-L3*H)*D1
340 NEXT I
```

Figure 9.31

```
RBC35

TYPE THE INFLUX TRANSFER COEFFICIENTS  K1,  K2,  K3
?0, .000068, .00039

TYPE THE LOSS RATES  L1,  L2,  L3
?.0028, .00014, .0013

TYPE THE INIT. VAL. FOR FOOD PEST. CONC., F1, F2
?31, 14

TYPE  H, THE INIT. CONC. OF PEST. IN THE HUMAN
?2


 J   S1(J)      N1(J)              N2(J)           N1(J)+N2(J)      H1(J)

 0  53.0001  31                 14              45              2
 1  46.2471  11.1401            14.5024         25.6425         4.97521
 2  39.4942  4.00331            14.8164         18.8197         5.51874
 3  32.7412  1.43863            14.9513         16.3899         5.39536
 4  25.9888   .516983           14.9161         15.4331         5.14877

THE VALUE OF THE CLOSENESS CRITERIA IS 25.6594
```

Typical Result From Program Shown in Figure 9.31

Figure 9.32

the days. Instructions 250 to 340 of the program calculate the increase in the concentrations during a single year using daily increments. Instructions 350 to 410 store the yearly concentrations and the loop specified in lines 250 to 340 is repeated 4 times to give 4 sets of yearly concentrations. A summary of these results is printed out by line 530 to 590. The experimentally determined yearly concentrations, E3(I), of the pesticide in the food are read in by statements 705-720. Each of these terms is the sum of the DDT and DDE concentrations of insecticide. The values, E4(I), of the concentration of the pesticide in the humans are read in by lines 730-745. These sets of values are used in the sum of squares closeness criteria, statement 830. In this statement, the expression (N1(W)+N2(W)-E3(W)) represents the deviation of the sum of the calculated concentrations of the pesticide from the actual concentrations of the pesticide in the food. The term (H1(W)-E4(W)) represents the difference between the calculated and the experimental values of the pesticide concentration in the human. These terms are each squared and added, in a loop, in lines 810 to 840 to give the value of the closeness criteria. The result is printed out by line 855.

The initial concentrations used in the run whose results are shown in figure 9.32 were taken from the work of O'Neill and Burke (1971). They used an analogue computer to carry out an "intuitive search" to obtain the "best" set of values for the transfer coefficients. Their paper did not specify what closeness criteria they used to evaluate the best set of parameters. As the parameter values were altered and the results of the runs compared, one set with another set, the insight and intuition of the investigators was increased. Such interaction between man and the machine is sometimes the only way of obtaining a set of values for use as starting values in an automated search routine. Most automated search routines require starting values quite close to actual values if the routine is to converge.

The ability to visually display the results on a graphic display terminal or on a plotting device is of great assistance in comparing the results obtained from different sets of parameters. There are also technicolor graphic units which permit the display of results in different colors. Such multicolor displays better enable the investigator to compare results. Three-dimensional representations of the results are another technique that is quite useful in analyzing computer output.

Because it is usually the case that so much data is being generated by a computer during the course of its solving a problem, it is frequently quite difficult to thoroughly appreciate the final results that are printed out. Consequently, a variety of combinations of results may have to be printed out. The selection of which combination of results to examine is not easy. The problem is made more difficult by the fact that computer time costs money as does the investigator's time. In addition, a significant amount of reprogramming may be necessary to capture and print the desired information. Thus, a judicious choice must be made to ascertain just what combinations of results are to be printed out. It is indeed enigmatic that the investigator using a computer based method of analysis can find himself deluged with results which he cannot fully appreciate and on the other hand the investigator limiting himself to pencil and paper type mathematics is frequently able to obtain only a very small fraction of his necessary results.

Relation of this Formulation to the Differential Equation Formulation

The usual formulation of compartmental models is expressed in the language of differential equations. In such a formulation, it is assumed that the intensities are varying continuously. Crudely speaking, this means that the intensity in a compartment measured at time I is at most only slightly different in magnitude than when measured an instant of time later, $I+\delta t$. Here the symbol δt designates a very small increment of time, i.e. an "instant" of time. Such a continuous variation implies that there are no sudden large changes in the quantities during a small time increment; there are only small changes in small time increments. In fact, by making the time increment small enough, the magnitudes of the changes can be made as small as desired. Because of this assumed behavior of the time variation of the intensity, it is appropriate to use the differential calculus notion of the instantaneous time rate of change of a variable. The instantaneous time rate of change of a variable is the limiting value of the quotient of the change in the variable and the time increment during which the change takes place as the time increment becomes arbitrarily small. Thus, the expression of the fundamental equations in the language of the differential calculus requires that Fick's principle be stated in terms of instantaneous rates of change (usually the word instantaneous is omitted, it being understood) of the concentrations. Loosely speaking, Fick's principle is stated as, the time rate of change of the concentration is the rate at which the concentration is being increased minus the rate at which it is being decreased.

In contrast, our formulation of Fick's principle stated that the magnitude of the change in the concentration during a time period was equal to the increase in concentration during that time period minus the loss or decrease in concentration during the same period. Thus, our formulation involved equations relating the magnitude of changes of variables during a finite time period.

Such equations are sometimes called finite difference equations. It can be shown that, as the time increment becomes vanishingly small, our formulation is equivalent to the usual differential equation formulation. Such a demonstration will not be made here as it is more appropriate for a mathematical treatment.

It is perhaps worthwhile to give some examples of the two formulations. The first example has already been presented in equation 9.7. The symbol

$$\frac{dQ}{dt}$$

denotes the time rate of change of the variable Q. Similarly, the symbol

$$\frac{dP}{dt}$$

denotes the time rate of change of the variable P. The differential equation formulation of the second problem would appear as

$$\frac{dQ1}{dt} = F1 - K1 \cdot Q1 + K2 \cdot Q2 - K3 \cdot Q1 \tag{9.9a}$$

and

$$\frac{dQ2}{dt} = K1 \cdot Q1 - K2 \cdot Q2 \tag{9.10a}$$

These two equations correspond to equations 9.9 and 9.10 respectively.

A comparison of the two forms of the respective equations shows that

`(Q1(I+1)-Q1(I))/D1`

and

`(Q2(I+1)-Q2(I))/D1`

are respectively analogous to the quantities

$$\frac{dQ1}{dt} \quad \text{and} \quad \frac{dQ2}{dt}$$

Note the comment in the line below 9.10 on page 9.16.

The differential equation formulation of the third example is

$$\frac{dQ1}{dt} = -K1 \cdot Q1 - K2 \cdot Q1 \qquad (9.11a)$$

$$\frac{dQ2}{dt} = K2 \cdot Q1 - K3 \cdot Q2 \qquad (9.12a)$$

and

$$\frac{dQ3}{dt} = K3 \cdot Q2 \tag{9.13a}$$

As a fourth and final example, the usual mathematical equivalents of equations 9.19, 9.20 and 9.22 are

$$\frac{dS}{dt} = -L1 \cdot S \tag{9.19a}$$

$$\frac{dP}{dt} = K2 \cdot S - L2 \cdot P \tag{9.20a}$$

$$\frac{dR}{dt} = K5 \cdot P - L3 \cdot R \tag{9.22a}$$

A comparison of these two formulations should reveal to the student how to translate from one form to the other. It is frequently the case that the literature will employ the notation that a dot (·) over a variable signifies the time rate of change of the variable. In this event, the above equations would appear as

$$\dot{S} = -L1 \cdot S \tag{9.19b}$$

$$\dot{P} = K2 \cdot S - L2 \cdot P \tag{9.20b}$$

$$\dot{R} = K5 \cdot P - L3 \cdot R \tag{9.22b}$$

Much work has been done on the analysis of multicompartment models. In the literature describing such work, it is usual to list an array or matrix of transfer coefficients and to then supply a brief discussion or list of rules as to the use of the table in setting up the governing equations. An example of this is the work of O'Neill, 1971. His work consists solely in the listing of transfer matrices corresponding to different ecological compartmental models. In all such work, it is assumed that the equations are linear; that is, only the first powers of the variables appear in the equations and no products of variables appear.

To illustrate the relation to our formulation, we begin by referring to the general 3 compartment model described in Fig. 9.14. For ease of discussion it will be assumed that there are no entering flows and the only exiting flows go from one compartment to another. Thus, F1=F2=F3=0 and E1=E2=E3=0. The equations describing the system may then be written as

$$Q1(I+1)-Q1(I)=(-(K1+T1)*Q1(I)+T2*Q2(I)+K3*Q3(I))*D1$$

$$Q2(I+1)-Q2(I)=(K1*Q1(I)-(K2+T2)*Q2(I)+T3*Q3(I))*D1$$

$$Q3(I+1)-Q3(I)=(T1*Q1(I)+K2*Q2(I)-(K3+T3)*Q3(I))*D1$$

Now, by recalling the derivation of these equations, it is seen that the term T2*Q2(I) is the flow from the second compartment to the first compartment and the term K3*Q3(I) is the flow from the third compartment to the first compartment. This suggests that a change in notation for the transfer coefficients T2 and K3 would be of assistance in accounting for the various terms in the equation. The double subscript notation of BASIC suggests itself. Thus, we set T(1,2)=T2 and T(1,3)=K3. The first equation may then be written as

$$Q1(I+1)-Q1(I)=(-(K1+T1)*Q1(I)+T(1,2)*Q2(I)+T(1,3)*Q3(I))*$$

In an analogous manner, we also introduce the further obvious notation: T(2,1)=K1, T(2,3)=T3, T(3,1)=T1 and T(3,2)=K2. The second and third equations may then be written as

$$Q2(I+1)-Q2(I)=(T(2,1)*Q1(I)-(K2+T2)*Q2(I)+T(2,3)*Q3(I))*D1$$

$$Q3(I+1)-Q3(I)=(T(3,1)*Q1(I)+T(3,2)*Q2(I)-(K3+T3)*Q3(I))*$$

Since the terms K1+T1, K2+T2 and K3+T3 are effective transfer coefficients for the flows leaving the first, second and third compartments respectively, it is natural to designate these transfer coefficients by T(1,1), T(2,2) and T(3,3) respectively. Thus, the equations may now be written in the form:

$$Q1(I+1)-Q1(I)=(-T(1,1)*Q1(I)+T(1,2)*Q2(I)+T(1,3)*Q3(I))*D1$$

$$Q2(I+1)-Q2(I)=(T(2,1)*Q1(I)-T(2,2)*Q2(I)+T(2,3)*Q3(I)*D1 \qquad (9.36)$$

$$Q3(I+1)-Q3(I)=(T(3,1)*Q1(I)+T(3,2)*Q2(I)-T(3,3)*Q3(I))*D1$$

An examination of the right hand sides of the above set of equations suggests that if the transfer coefficients are written in the square array form

$$\begin{pmatrix} -T(1,1) & T(1,2) & T(1,3) \\ T(2,1) & -T(2,2) & T(2,3) \\ T(3,1) & T(3,2) & -T(3,3) \end{pmatrix}$$

and the concentrations in the compartments written in the vertical single element array form

$$\begin{pmatrix} Q1(I)*D1 \\ Q2(I)*D1 \\ Q3(I)*D1 \end{pmatrix}$$

that there is a simple rule for forming any of the right hand sides of equation (9.36) with the aid of these two arrays. The rule is the following: To form the right hand side of any equation multiply the terms in the corresponding row of the square array in order by the terms in the vertical array and add the results. For example, to obtain the right hand side of the second equation, multiply the 1^{st} element in the second row of the square array by the 1^{st} element in the single vertical array, then multiply the second element in the second row of the square array by the second element in the vertical array and finally multiply the third element in the second row of the square array by the third element in the single vertical array. The sum of these products then gives the right hand side of the second equation.

The student will note that the indices of a term in the square array specify the row and the column of the transfer coefficient. Thus, T(2,3) is the element in the second row and the third column. In general, T(I,J) denotes the transfer coefficient in the I^{th} row and the J^{th} column. Furthermore, the term T(2,3) designates the proportion of the substance in the third compartment transferred in the time increment D1 to the second compartment. Hence, in general, the term T(I,J) designates the proportion of the substance of the J^{th} compartment transferred in the time increment D1 to the I^{th} compartment.

In mathematics it is usual to designate the square array by T(I,J) or simply T. If Q(J), or simply Q, designates the vertical single array, the right hand side of equation 9.36 may be written as

$$T \cdot Q \quad .$$

It is usual practice to also designate the source and leakage terms,

$$\begin{pmatrix} F1 \\ F2 \\ F3 \end{pmatrix} \quad \text{and} \quad \begin{pmatrix} E1 \\ E2 \\ E3 \end{pmatrix}$$

by F(J) and E(J) or more simply by F and E respectively. By using this notation, the description of the original model depicted in figure 9.14 can be written in the succinct form

$$Q(I+1) = Q(I) + T \cdot Q + F - E.$$

The expression of the equations is such an abbreviated form it is sometimes helpful in obtaining an overall description of the behavior of the system. Furthermore, it does permit the elaborate machinery developed in the matrix calculus to be used to assist in the analysis of the system. Nevertheless, when actual numerical values are desired, as they most usually are, the original form of the equations must be used. In this way the matrix notation can effectively hide the existence of the necessity of performing a "grubby lot" of arithmetic.

The student who is familiar with linear algebra or with the use of the MAT commands in the BASIC programming language will have long recognized that the square array is a 3x3 square matrix and the vertical linear array is a 3 vector. We apologize to these students for this digression, nevertheless, it has been your author's experience that many students are unfamiliar with, or do not recall, matrix operations and their use. Consequently, this brief discussion was included in the work.

Such a matrix is naturally called a transfer matrix. Since the above described use of the matrix to form the fundamental equations permits a ready and obvious extension to the description of a system consisting of more than three compartments, the notion of a transfer matrix and its properties plays a central role in the analysis of multicompartment models. As stated previously, O'Neill (1971) has presented several transfer matrices, each of which is a listing of the transfer coefficients corresponding to a particular compartment model. He also provides the reference to the original investigator in order that the interested reader can more easily use the particular matrix in the investigator's program.

If the system under examination is in the steady state, the amount of flow entering any compartment in an increment of time is balanced by the amount of flow leaving the compartment in the same interval of time. This implies that, in the transfer matrix, the diagonal term is the negative of the sum of all of the rest of the terms in the same column. For example, the diagonal term, $-T(3,3)$, is such that

$$-(T(1,3) + T(2,3)) = -T(3,3).$$

This condition holds for all transfer coefficient matrices of interest since the determination of the transfer coefficients is usually accomplished when the system being modeled is in the steady state.

PROBLEMS

CHAPTER IX

1. Using equation (9.5), discuss the constant infusion of ions into the tissue. Discuss the results for different infusion rates and different transfer coefficients. For reference see Chapter 3, Atkins.
2. Using equation (9.5), analyze the effects of multiple infusions of a tracer labeled substance. In this case, the tracer is added in equal doses at equal time intervals.
3. Using equation (9.5), discuss the case corresponding to the infusion of a tracer labeled substance whose rate of infusion decreases in proportion to the elapsed time.
4. Write a computer program to analyze the formation and excretion of paracetamol sulphate from paracetamol. Use the transfer coefficients K1 = 0.09 per hour, K2 = 0.753 per hour, and K3 = 0.257 per hour. Assume a unit initial concentration of paracetamol in the plasma. These values are taken from Atkins (1969). How does a change in the time step D1 affect the results?
5. In the two compartment fallout problem, run the program for different intake and loss rates. What effect do these variations have on the results? How do the results change as the initial value of the intensity of the source changes?

6. Consider the lichen-caribou-Eskimo food chain example. Alter the program for determining the parameter A to use a time increment of one-tenth of a day. How much does the value of A and F(0) change? As a result of this alteration, what can you say about the sensitivity of the final values to changes in the length of the time increment?

7. Write a program to determine the value of the decay parameter when the half life is specified. Choose some different half lifes and determine the decay parameter by trial and error. What effect does changing the value of the time increment have on the final values of the decay parameters? What is the relation of the time unit in which the decay parameter is expressed to the length of the time increment, D1?

8. (a) Write a program to calculate the time variation of the radiation intensity on the lichen-caribou-Eskimo food chain. Run the program for different values of the parameters. Discuss the results.

   (b) Alter the preceding programs to include the time variation of the lichen diet of the caribou. Use the variations as specified in the text. Try other variations. Compare and discuss the results.

   (c) Alter the program in part (a) to include the time variation of the caribou diet of the Eskimo. Use the time variation specified in the text as well as your own. Compare and discuss the results.

(d) Combine parts a, b, and c. Compare and discuss the results. NOTE: This problem could be parceled out to the class for a total class project.

9. Using the program describing the movement of pesticides through a food chain, figure 9.31, choose different transfer coefficients and loss rates and make a set of runs. Compare the results.

10. In the program listed in figure 9.31, the measure of closeness is the sum of the squares of the deviations of the calculated and the experimental data. The terms of this sum are calculated by line 830. Alter this instruction so that the measure of closeness is the sum of the squares of the relative errors. Using the same set of transfer coefficients and loss rates as you used in problem 9, rerun the program. Compare your results. Can you improve the measure of closeness?

11. Alter the measure of closeness in one of the programs to be the sum of the squares of the relative errors or the minimum of the magnitude of the largest deviation. Make some runs and compare the results.

REFERENCES

CHAPTER IX

Atkins, G. L., 1969. Multicompartment Models for Biological Systems. Methuen & Co. Ltd., London.

Born, G. V. R. and Bulbring, E. 1956. The Movement of Potassium Between Smooth Muscle and the Surrounding Fluid. J. Physiology, 131, 690.

Cummings, A. J., King, M. L. and Martin, B. K., 1967. A Kinetic Study of Drug Elimination: The Excretion of Paracetamol and its Metabolites in Man. Br: J. Pharmaceutical Chemotherapy, 29, 150.

Eberhardt, L. L., 1970. Notes on an Introduction to Food Chain Kinetics. Work performed under A.E.L. contract AT(45-1)-1830.

Eberhardt, L. L. and Hanson, W. C., 1969. A Simulation Model for an Arctic Food Chain. Health Physics, Vol. 17, pp. 793-806.

Hanson, W. C. and Eberhardt, L. L., 1969. Effective Half-Times of Radionuclides in Alaska Lichens and Eskimos: In Symposium on Radioecology. Edited by D. J. Nelson and F. C. Evans; Conf. -670503 U.S.A.E.C., Washington, D. C.

Hevesey, G. and Hahn, L., 1940. Turnover of Lecithin, Cephalin and Sphyngomyelin. Biol. Meddr. 15(5).

Monot, C. and Martin, J. (1974). Reflections on Some Algorithms for the Solution of the Inverse Problem (Identification or Adjustment) for Linear Compartment Models. In Mathematical Models in Biology and Medicine. Edited by Bailey, N. T. J., Sendov, Bi. and Tsanev, R. p. 49.

O'Neill, R. V. and Burke, O. W. (1971). A Simple Systems Model for DDT and DDE Movement in the Human Food Chain. ORNL-IBP-71-9. Oak Ridge National Laboratory, Oak Ridge, Tennessee.

Popjak, G. and Beeckmans, M. L., 1950. Synthesis of Chloresterol and Fatty Acids in Foctuses and in Mammary Glands of Pregnant Rabbits. Biochemical Journal. 46, 547.

Riggs, D. S., 1967. The Mathematical Approach to Physiological Problems, A Critical Primer. The M.I.T. Press, Cambridge, Mass.

Rittenberg, D., 1951. A Method for the Evaluation of the Rate of Protein Synthesis in Man. In Ciba Foundation Conference, Isotopes in Biochemistry, p. 190.

Sheppard, C. W., 1962. Basic Principles of the Tracer Method. Introduction to Mathematical Tracer Kinetics. J. Wiley & Sons, Inc., New York, NY

Sprinson, D. B. and Rittenberg, D., 1949. The Rate of Interaction of the Amino Acids of the Diet with the Tissue Proteins. J. Biol. Chem. 180, 715.

Zender, R., Denkinger, E. and Falbriard, A., 1965. Dicroissance Plasmatique de Substances Glomerulaires chez le Lapin. Helvitia Physiological Pharmacy Acta. 23, 199.

SIMULATING TREE GROWTH WITH A COMPUTER

by

W. R. Pierce

School of Forestry
University of Montana
Missoula, Montana

ABSTRACT

Using a computer to simulate tree growth allows one to quickly see how stocking level affects tree diameter, height, and volume growth, as well as stand mortality and volume. Also, one can easily see how individual tree growth is influenced by tree size and age. While this is not a new or unique concept in forestry, the approach presented here differs from that of such authors as Beck (1974), Lin (1974), and Burkhart and Strub (1974), who have used mathematical formulas to simulate natural growth phenomena.

The ability of a computer to do many calculations in a short period of time makes the substitution of a computer language for formal mathematics possible. This paper will describe a method of using the computer directly to simulate growth of ponderosa pine and Douglas-fir in stands of various stocking levels. Using the computer directly offers several advantages over constructing mathematical formulations: (1) it is faster, (2) it allows use of BASIC computer language, which is easy to learn, (3) it clearly illustrates how growth rate changes over the life of a tree or a stand, and (4) it yields results that can be understood by a wider audience.

The purpose of these notes is to describe to the student how a tree grows in terms that provide a visual representation or feel of the presentation. As trees increase in size; some of the dimensional changes are great enough to be observed from year to year or even from month to month. As a tree grows in height the change is easily observed as the growing season progresses.

The year to year changes in diameter are more subtle but can be detected by the use of "dendrometer" bands around the tree with an expansion scale indicating the dimensional changes. The third characteristic of a tree, and the most important, is the stem volume. Year to year changes of this dimension are not easily depicted. To add to the confusion, the volume is not easily measured or computed.

The stem contains all of the usable material for today's markets so our efforts at depicting growth will include this dimension. It should be remembered, however, that the stem is only a portion of the biomass of a tree and that parts which you do not use through consumption are eventually oxidized and their constituents returned to the soil and air. Since most of the minerals and nutrients that a tree removes from the soil during its growth are in these none used parts their return whence they came is of great importance and until they oxidize they are a source of stored energy.

The stem of a tree is a complicated geometric figure and the computation of its cubic volume is, at its best, only an approximation. The most accurate method is accomplished through measuring its displacement when immersed in a fluid. But this kind of treatment is not compatible with a standing growing trees, so many methods of approximation have been developed. Probably the most accurate of these, consistent with ease and speed of application, are local volume tables which require only the measurement of a tree's diameter for their application. The most easy to apply and the most accurate of local volume tables are the TARIF volume tables. These are a series of volume tables for different TARIF numbers. The TARIF number of a tree can be determined from a table of these numbers if you know the diameter and height of a tree. More explicitly, the TARIF number is the amount of cubic volume from a 12 inch stump to a 4 inch top to one square foot of basal area at a point 4.5 feet from the ground.

The cubic volume itself is computed from a multiple regression formula. All of this information does not give the reader much feel for how a tree changes in size with increases in age. An attempt is made here to express these changes in more readily understandable relationships and in simple non-mathematical language.

Custom has decreed that a tree's diameter will be measured at a point 4.5 feet above the average ground level. It takes a seedling a period of time to reach this height, for our illustration we will use 7 years. After the 7th year then a tree begins to accumulate a diameter dimension. Each year sees a growth ring added to the accumulation of previous years growth. The amount that is added varies with tree species, tree size, the productivity of the growing site, the amount of moisture available during the growing season, and the general physical condition of the tree.

For our illustration we will use ponderosa pine as the tree species and a productive site index of 100. The index is the height of a tree at 100 years. This site index is about average or in the middle of the range between a very poor site and a site of maximum productive capacity. At the close of the 8th year the diameter of the tree should be the amount of one year of growth. The value selected for this diameter growth is .28 inches Individual trees may grow faster than this or they may grow less. This value is the diameter growth rate of the average tree depicted in the "Yield of Even-aged Stands of Ponderosa Pine" by Walter Meyer. (Meyer, 38).

The diameter growth rate will become less as the tree gets larger. A constant diameter growth results in a rapidly increasing volume, since the size of the tree upon which this diameter is laid is steadily growing larger. As the tree gets larger, both the diamter growth rate and cross sectional area growth rate of the tree stem steadily decrease. The cross sectional area rate of growth is about double that of the diameter growth The reduction in diameter growth is small at first and then increases as the tree gets larger. The diameter growth seems to be related more to tree size than age, since quite old trees, if small, have the ability to put on diameter growth equivalent to younger trees of the same size.

Since the reduction in diameter growth is a function of tree size, it can be handled as follows:

```
10  DIMENSION D(201)
20  D1=.282
30  D4=.00228
40  FOR I=8 TO 200
50  D(I+1)=D(I)+D1-D4*D(I)
60  NEXT I
70  END
```

The value of D4 can be determined from growth records of trees. There will be more about a selection of this value and Dl later.

Besides growing in diameter trees increase in height each year and this will now be added to the tree growth model. It has already indicated that the young tree reaches breast height, 4.5 feet at about 7 years of age so the annual heighth increase will be started at this point as was the diameter growth.

Like diameter growth, the heighth growth becomes less each year as the tree becomes larger but in this case the effective dimension is height. The program will be changed to include this growth parameter of a tree, but to make the handling of variables a little easier a new variable will be used in place of B.

The program will now look as follows:

```
10  DIM P(201,5)
20  Dl=.282
30  D4=.00228
40  D2=2.3
50  D5=.0193
60  P(8,2)=4.5
100 FOR I=8 TO 200
140 IF P(I,2)<100 THEN 200
150 D5=.0152
200 P(I+1,1)=P(I,1)+Dl-D4*P(I,1)
210 P(I+1,2)=P(I,2)+D2-D5*P(I,2)
270 NEXT I
280 END
```

The variables are defined as follows:

1. D1=diameter growth of the tree in inches
2. D2=height growth of the tree in feet
3. D4=a constant of proportionality that reflects the loss of diameter growth because of the size of the diameter
4. D5=a constant of proportionality that reflects the loss of height growth because of the height of the tree
5. I=age of the tree in years
6. P(I,1)=diameter of tree in inches in the Ith year
7. P(I,2)=height of tree in feet in the Ith year

A free growing tree will have a root system and crown that is in direct proportion to the stem size. No attempt will be made here to include these dimensions in the program but the volume of the stem will now be included. Tarif numbers which provide the volume table to be used on a given tree are determined from a table of entry numbers using the diameter and height of the tree. Once the tarif table has been determined the volume of the tree is then determined by its diameter.

To provide for volume in the program, the following statements will now be added:

```
 90  V2=.0006
110  IF P(I,1)<2 THEN 140
120  V=.26
130  V1=V1+P(I,1)/P(I,2)
```

```
160  IF I<60 THEN 190
170  IF I>135 THEN 190
180  V2=V2-.0000055
190  V=V+V2*I
220  P(I,5)=V*V1*P(I,1)
```

New variables added to the program are:

V and V1=constants of proportionality whose product provides the total cubic volume of the tree in feet when multiplied by the diameter of the tree

V2=constant of proportionality that provides an increase in V as the age of a tree increases

P(I,5)=total volume of the tree in cubic feet in the Ith year

The minimum diameter tree for which volume is computed is 2", the value of V is not assigned until the tree reaches this size. At the same time the value of V1 begins to accumulate with the yearly increase equal to the ratio of the tree diameter in inches over the height in feet.

The program at this point will grow a free standing tree without environmental limitations. The tree grows by increasing its diameter and height, both values increasing by adding on a yearly growth. The yearly growth reduces in amount each year as the tree gets larger.

The volume in turn is computed by applying the product of 2 values both of which increase as the tree becomes older. Volume of the stem of a tree is a function of more than diameter and height, it also is affected by the rate of taper of the stem, this taper is rapid in young trees, gradually lessening as the tree ages. Program statement 130 provides for this improvement in form. When the tree is about 60 years old, the rate the form improves begins to drop off. Statement 180 provid for this change which continues until the tree is about 135 years old.

The program needs statements that will PRINT the data it has computed so the following are added:

```
 70  PRINT "AGE   DBH   HEIGHT   VOLUME"
 80  N=10
230  IF I<>N THEN 270
240  PRINT I;TAB(6);INT(P(I,1)*100+.5)/100;TAB(13);
250  PRINT INT(P(I,2)+.5);TAB(23);INT(P(I,5)*10+.5)/10
260  N=N+10
```

A RUN of this program (figure 2) provides the information listed in figure 1.

```
AGE     DBH     HEIGHT     VOLUME
10      0.56     9           0
20      3.34     28          0.5
30      6.05     44          3.1
40      8.71     58          8.1
50      11.31    69          16
60      13.84    78          27
70      16.32    85          41.1
80      18.74    91          58.5
90      21.11    96          79.4
100     23.43    100         103.6
110     25.69    107         130.7
120     27.9     114         160.1
130     30.06    119         191.4
140     32.18    124         227.2
150     34.24    127         269.4
160     36.26    131         315.7
170     38.23    134         366.3
180     40.16    136         421.4
190     42.05    138         481
200     43.89    140         545.4
```

Figure 1

Growth table for ponderosa pine on site index 100

```
10 DIM P(201,5)
20 D1=.232
30 D4=.00223
40 D2=2.3
50 D5=.0193
60 P(8,2)=4.5
70 PRINT" AGE    DBH    HEIGHT    VOLUME"
80 N=10
90 V2=.0006
100 FOR I=8 TO 200
110 IF P(I,1)<2 THEN 140
120 V=.26
130 V1=V1+P(I,1)/P(I,2)
140 IF P(I,2)<100 THEN 160
150 D5=.0152
160 IF I<60 THEN 190
170 IF I>135 THEN 190
180 V2=V2-.0000055
190 V=V+V2*I
200 P(I+1,1)=P(I,1)+D1-D4*P(I,1)
210 P(I+1,2)=P(I,2)+D2-D5*P(I,2)
220 P(I,5)=V*V1*P(I,1)
230 IF I<>N THEN 270
240 PRINT I;TAB(6);INT(P(I,1)*100+.5)/100;TAB(13);INT(P(I,2)+.5);
250 PRINT TAB(23);INT(P(I,5)*10+.5)/10
260 N=N+10
270 NEXT I
280 END
```

Figure 2

A basic language program

for growing a ponderosa pine on site index 100

The volumes produced by this program for a tree as it increases in diameter and height is a product of two factors and the diameter of the tree. One of these factors is an accumulating ratio of the diameter of the tree over the height for each year in the life of the tree. This value accounts for the improving form of the tree with age, the other is V which increases by the product of V2 times the age class each year.

The resulting values are lower than the volumes produced by a tarif volume table. This should be expected because the entry tables used to determine the tarif number for a tree of a given diameter and height are constructed from the calculations of the volumes of many ponderosa pines for each combination of diameter and height. The measurements used in the calculations were taken of trees grown in a forest, of trees growing in competition with other trees.

The volumes of trees growing alone, without competition were not a part of the measurements. Such trees retain their branches as part of their live crown for most of their total height. The shortest or greatest taper in a tree stem is in the crown and the most rapid diameter growth rate is at the base of the crown. This results in a tree of low form factor or lower volume for a given diameter and height. The variation of the computed volume from the tabular volume increases as the tree becomes older and larger.

The history of development of open grown ponderosa pine in a forest environment is not available, even for a small number of trees. Those individuals of this species grown in fields or parks certainly would not be representative of forest grown trees where the only change has been a lack of competition from other trees.

Without some basic information to use for comparison there is no way to judge whether the diameters produced by running this program are realistic. The constants D1 and D4 cannot be determined by testing against known diameters for different ages for the open grown pine but they can be determined by fitting computed diameters to the diameters provided in the yield tables for fully stocked even aged stands of this species. These tables are not the history of the development of individu stands but are a composite of many stands of different ages put together to illustrate the historical development of a single stand. They are the best information available about the growth in size, number, and volume of trees in well stocked stands.

Height growth of trees is not materially affected by competition from other trees in even aged stands so the height growth of this individual tree matches that of the average tree in the yield table.

Since the computed diameter and volume have no information for comparison and the constants producing these values need basis in fact our next step is to build on our program so that it can duplicate a yield tabl then our constants can be adjusted to accommodate known trends.

Yield tables represent the so called normally stocked stand or a forest that completely utilizes the site on which it grows. This criteria of stocking is subject to some question but this should not affect their use to determine the growth variables required for the growth model. Only one variable is added to the first model but this is a variable that is very influential to the growth rate of a tree. There is now competition for the nutrients and moisture that a tree needs to grow and there is competition for its required space and light. The above ground competition changes the shape of the tree. Lower limbs, not receiving enough light die and eventually drop. This leaves less crown to provide the energy required for growth. In all, the tree grows less rapidly as competition increases and competition increases as the individual trees grow and require more space for their activities.

How many seedlings start a forest on its growth cycle? Yield tables do not provide this information so an arbitrary number of 2300 per acre is selected. During the first two or three years the mortality would be quite severe, but they soon outgrow the early hazards and the survival rate settles down to a predictable amount. No attempt will be made to predict these early year mortality patterns, instead a mortality, variable D, will be sought that reduces 2300 trees per acre to the numbers of trees per acre that the yield table lists by 10 year age classes starting with the age of 20.

A mortality of 3.04% per year reduces 2300 trees to the required number, 1280 at age 20. At 20 years of age the mortality in the program changes to 2.45 percent. When the basal area exceeds 75 square feet the mortality is increased by the addition of the product of .000015 times the basal area in excess of 80 square feet. At the maximum of 230 square feet this amount to .000015*(230-80)=.00225 or it raises the mortality to 2.675 percent, an increase of .225 percent over the mortality of trees without any reduction in their ability to survive in competition with other trees.

When the age of the trees passes 45 years the mortality is increased by the rate of 61.3/45. Each year thereafter the mortality is reduced as the tree becomes older. The size of the variables that apply the effect of age and stocking on the mortality rate are determined through trial and error. They were selected to make the number o trees surviving each decade approximate those of the yield table.

These changes can be made in the program by adding the following statements.

```
 14  P(1,3)=2300
 15  D=.0304
 16  D6=56.5
115  P(I,4)=(P(I,1)/2)**2*3.142/144*P(I,3)
131  IF I<8 THEN 140
132  P(8,2)=4.5
192  IF I<21 THEN 197
193  D=.0245
194  IF P(I,4)<80 THEN 197
195  D0=.000015
196  GO TO 198
197  D0=0
198  D8=D+D0*(P(I,4)-80)
199  IF I<60 THEN 204
200  D8=D8*D6/I
204  IF I<8 THEN 220
```

and changing the number of statement 200 to 205.

Statement 60 can now be removed and 70 and 100 can be changed as follows:

```
       70  PRINT " AGE   DBH   HEIGHT   NO. OF TREES BASAL AREA
             TOTAL VOL."
      100  FOR I=1 TO 200
```

Statement 250 should now read

```
      250  PRINT TAB(25);INT(P(I,3)*10+.5)/10;TAB(35);INT(P(I,4)+.5);
      251  TAB(46);INT(P(I,5)+.5)
```

The new variables are:

D=mortality as a decimal

D0=a value that increases mortality. This value is multiplied by the basal area in excess of 80 square feet.

D6=a value that changes mortality when it is divided by age starting at 60 years and the ratio used as a multiplier for D.

D8=the corrected mortality.

P(I,3)=the number of trees per acre in the Ith year

P(I,4)=the basal area, in square feet, per acre.

The results of running this program would indicate that an optimum survival rate for the trees exists when they reach the maximum age computed for the yield table. When the trees are 200 years old, their mortality would be .0245+.000015*(228-80)*56.5/200=.0075.

The lowest survival rate occurs at the age of 60 years when mortality reaches .0245+.000015*(227-80)=.026705.

It is during the age period from 20 to 60 that the basal area per acre is approaching its maximum value and the effect on mortality the rapid basal area per acre increase over the previous years reaches its culmination. From 60 years on, the mortality starts a steady decrease with the increasing age of the trees.

Various techniques have been proposed and used in tree growth models to account for the effect of other trees on the growth rate of a individual. There are so many factors that effect the growth of individual trees, but are very difficult to measure or control, that it is an exercise in futility to develop methods of accounting for the effect on growth of a selected tree caused by other trees. Much resear is required before realistic models can be constructed of individual tree growth in a forest. It is more logical to grow the average tree in a forest and for this purpose the basal area of a stand is the best criteria of competition available (Beck, 74). Yield tables provide a measurement of this parameter in their tables making its use relatively easy.

The next step is to add the statements that will apply these basal area values to the diameter growth to achieve a duplication of their real effect.

Competition as measured by basal area appears to start reducing the diameter growth of a tree when it reaches 130 square feet for up to that point an annual diameter growth of D1=.282 inches that reduces each year by D4=.00228 times diameter (P(I,1)) fits very well to the pattern of diameter growth depicted by the yield table. This trend of diameter growth continues until the basal area reaches 130 square feet per acre, or an age of about 25 years but beyond this point, computed diameters begin an increasing gain over the yield table diameters.

The addition of a statement identifying this basal area:

```
151  IF P(I,4)<130 THEN 155
```

and the addition of statements to account for the effect of basal area on diameter growth and to change the variable for annual diameter growth from D1 to D9.

```
152  D9=D1-D3*P(I,4)
153  GO TO 160
155  D9=D1
```

and a statement that assigns a value to D3;

```
65  D3=.0006
```

and to change statement 140 so that it branches to 151;

```
140  IF P(I,2)<100 THEN 151
```

and to change statement 205 to substitute D9 for D1

```
205  P(I+1,1)=P(I,1)+D9-D4*P(I,1)
```

will provide for the effect of basal area per acre on the diameter growth of the average tree.

The variable D3 is a constant that, when multiplied by the basal area when it reaches a minimum of 130 square feet, will further reduce the diameter growth of D1=.282. The value of D3 like the other constants used in this model has been deliberately selected to produce results that duplicate the yield table values.

Now go back to the statement that calculates the volume of the tree and include the number of trees on an acre to calculate the volume per acre.

```
220  P(I,3)=V*V1*P(I,1)*P(I, 3)
```

The variable V=.26 which is increased by an amount of, V2=.0006 times the age for every year of growth beyond seven years and V1, which is an accummulated ratio of diameter over height for each year in the life of the trees after the diameter has reached 2 inches, were determined by comparing the volumes that resulted for the individual tree, computed by taking the product of V*V1*P(I,1) with that obtained from a TARIF volume table for a ponderosa pine of similar diameter. The tarif number that identifies the volume table is determined from an entry table that uses the generated diameters and height for a tree of that age.

The comparison was not very good, for the generated volumes exceed those indicated by the TARIF volume tables from the ages of 60 to 135 years. A correction to the program is made that reduces the value of

V2 a little each year from the age of 60 to 135. V, which represents the improvement in form of a tree as it grows older, is increased by adding V2 times age to V each year. During the years between 60 and 135 apparently the improvement in form is slower than during the balance of the trees life and a reduction of V2=.0006 by .0000055 each year provides the needed change.

The necessary statements are:

```
160 IF I<60 THEN 190
170 IF I>135 THEN 190
180 V2=V2-.0000055
```

These statements are already in the program placed there during the construction of the individual tree growth model but not explained at that time. Why this slow down in a tree form improvement occurs during this period cannot be determined for sure but it might be caused for the following reason. The years from 20 to 60 represent the period of rapid basal area increase and hence greatly increasing competition between trees. This causes the lower branches of a tree to die and, to eventually drop off. As indicated earlier, the maximum rate of taper of a tree is in the crown. As the crown becomes shorter in relation to the total height of a tree the form of the tree improves. From 60 years of age on, the basal area does not change much so the competition remains constant, with mortality complementing growth. This should result in a reduction in the die-off rate of the limbs at the bottom of the crowns of the trees, with a resulting slow down in the form or taper improvement of the trees as they continue to grow in height and diameter. This continued height growth still provides increasing shading of the lower limbs but there is no longer the increase of competition from the side so the crown length to total tree height is reduced to a very slow rate of change.

After a tree reaches an age of 135 its annual height growth has reduced from an initial rate of 2.3 - .0193*4.5=2.21 feet at seven years of age to a rate of 2.3 - .0152*121=0.46 feet. The next 65 years will result in only 19 more feet of height. At the start of th period the diameter growth of the tree is .282- .0006*230 - .002*18.3= inches and it is still growing at the rate of .28 - .0006*230-.002*24.7 .094 inches when the tree is 200 years old. This annual increase in diameter is fairly uniform all the way up the stem to the base of the crown. The result is a steady improvement in form that will continu so long as the tree is able to maintain its diameter growth rate.

The single tree growth program, which was produced first, utilizes all the constant that effect diameter, height and volume growth as determined from building the program to duplicate the yield table values for these dimensions for well stocked stands of ages from 20 to 190 years with the exception of those duplicating the effect of basal area on tree growth in diameter and the constants required to duplicate the number of trees per acre.

The changing of statement 14

14 P(1,4)=2300 to an INPUT statement

```
14   INPUT P(1,4)
```

lets you vary the number of trees per acre that the program uses for a run. To facilitate this input the following statement should also be added:

```
13    PRINT "HOW MANY TREES PER ACRE DO YOU WISH TO PLANT";
```

It should be remembered, however, that the reduction of this starting number of trees to the number of trees at 20 years of age may not be realistic. The comparisons should be made with number of trees surviving to the 20th year.

A listing of the completed program is given in figure 3 and followed by three sample run outputs in figures 4, 5, and 6. The first of these runs duplicates the yield table values and the other two show the results of beginning with less and then with more than 2300 seedlings per acre. Table 7 gives normal yield table values with tarif volumes.

```
10 DIM P(201,5)
13 PRINT"HOW MANY TREES PER ACRE DO YOU WISH TO PLANT";
14 INPUT P(I,3)
15 D=.3304
16 D6=56.5
20 D1=.282
30 D4=.00223
40 D2=2.3
50 D5=.0193
65 D3=.0036
70 PRINT" AGE   DBH   HEIGHT   NO. OF TREES BASAL AREA TOTAL VOL."
80 N=10
90 V2=.0006
100 FOR I=1 TO 200
110 IF P(I,1)<2 THEN 131
115 P(I,4)=(P(I,1)/2)**2*3.142/144*P(I,3)
120 V=.25
130 V1=V1+P(I,1)/P(I,2)
131 IF I<8 THEN 140
132 P(8,2)=4.5
140 IF P(I,2)<100 THEN 151
150 D5=.0152
151 IF P(I,4)<130 THEN 155
152 D9=D1-D3*P(I,4)
153 GO TO 160
155 D9=D1
160 IF I<60 THEN 190
170 IF I>135 THEN 190
180 V2=V2-.0000055
190 V=V+V2*I
192 IF I<21 THEN 197
193 D=.0245
194 IF P(I,4)<80 THEN 197
195 D0=.000315
196 GO TO 198
197 D0=0
198 D3=D+D0*(P(I,4)-30)
199 IF I<50 THEN 204
200 D3=D3*D6/I
204 IF I<3 THEN 220
205 P(I+1,1)=P(I,1)+D9-D4*P(I,1)
210 P(I+1,2)=P(I,2)+D2-D5*P(I,2)
220 P(I+1,3)=P(I,3)-D3*P(I,3)
225 P(I,5)=V*V1*P(I,1)*P(I,3)
230 IF I<>N THEN 270
240 PRINT I;TAB(5);INT(P(I,1)*100+.5)/100;TAB(13);INT(P(I,2)+.5);
250 PRINT TAB(25);INT(P(I,3)+.5);TAB(35);INT(P(I,4)+.5);
251 PRINT TAB(46); INT(P(I,5)+.5)
260 N=N+10
270 NEXT I
280 END
```

Figure 3

Basic computer language for generating
ponderosa pine yield tables for site index 100

```
HOW MANY TREES PER ACRE DO YOU WISH TO PLANT ?2300
 AGE    DBH    HEIGHT    NO. OF TREES BASAL AREA TOTAL VOL.
 10     0.55   9             1742        0           0
 20     3.34   28            1279        78          653
 30     5.61   44            935         169         2789
 40     7.14   58            756         210         4596
 50     8.47   69            577         225         6095
 60     9.72   78            440         227         7014
 70     10.95  85            347         227         7714
 80     12.14  91            283         228         8342
 90     13.3   96            236         228         8909
 100    14.44  100           201         229         9420
 110    15.54  107           174         229         9850
 120    16.62  114           152         229         10191
 130    17.68  119           135         230         10451
 140    18.7   124           120         230         10781
 150    19.71  127           108         230         11231
 160    20.69  131           98          229         11675
 170    21.66  134           90          229         12115
 180    22.6   136           82          229         12553
 190    23.53  138           76          228         12990
 200    24.43  140           70          228         13429
```

Figure 4

Ponderosa pine normal yield table
for a site index of 100

```
HOW MANY TREES PER ACRE DO YOU WISH TO PLANT ?1000
 AGE      DBH    HEIGHT    NO. OF TREES BASAL AREA TOTAL VOL.
 10      0.56   9              757          0            0
 20      3.34   28             556          34           234
 30      6.06   44             431          86           1341
 40      8.63   53             335          135          2701
 50      10.37  69             259          152          3731
 60      12.01  78             200          157          4482
 70      13.59  85             159          160          5030
 80      15.11  91             131          163          5617
 90      16.59  96             110          165          6104
 100     18.03  100            94           166          6543
 110     19.42  107            82           168          6920
 120     20.77  114            72           169          7227
 130     22.09  119            64           170          7470
 140     23.37  124            57           170          7753
 150     24.62  127            52           171          8129
 160     25.84  131            47           171          8493
 170     27.04  134            43           171          8853
 180     28.2   136            39           171          9209
 190     29.34  138            36           171          9564
 200     30.45  140            34           171          9918
```

Figure 5

Ponderosa pine yield table for a starting generation of 1000 trees per acre for a site index of 100

```
RUNNH


HOW MANY TREES PER ACRE DO YOU WISH TO PLANT ?10000
 AGE    DBH    HEIGHT   NO. OF TREES BASAL AREA TOTAL VOL.
 10     0.56    9          7574         0           0
 20     2.36    28         5562         243         2238
 30     3.32    44         4184         333         6396
 40     4.5     58         3135         346         9005
 50     5.15    69         2349         339         10543
 60     5.84    78         1763         328         11431
 70     6.57    85         1372         323         12130
 80     7.3     91         1106         321         12800
 90     8.02    96         914          321         13422
 100    8.72    100        772          320         13989
 110    9.42    107        662          320         14462
 120    10.09   114        576          320         14322
 130    10.76   119        506          320         15032
 140    11.41   124        450          319         15456
 150    12.05   127        403          319         16011
 160    12.68   131        363          318         16564
 170    13.29   134        330          318         17116
 180    13.9    136        301          317         17672
 190    14.5    138        276          316         18232
 200    15.03   140        254          316         18797
```

Figure 6

Ponderosa pine yield table for a starting regeneration of 10,000 trees per acre for a site index of 100

612

SITE INDEX 100, TECHNICAL BULLETIN NO. 630

AGE	DBH	HEIGHT	NO. OF TREES	BASAL AREA	TOTAL VOL.
20	3.6	30	1230	93	837
30	5.5	44	1000	165	3005
40	7.0	55	735	210	4934
50	8.5	65	574	226	6301
60	9.7	73	445	228	7139
70	10.9	80	352	228	7760
80	12.1	88	286	228	8299
90	13.3	94	236	228	8731
100	14.5	100	190	228	9095
110	15.6	106	172	228	9748
120	16.6	111	152	228	10371
130	17.7	116	134	228	10860
140	18.7	121	120	228	11235
150	19.7	125	108	228	11474
160	20.7	129	98	228	11814
170	21.7	133	89	228	12136
180	22.6	136	82	228	12186
190	23.5	139	76	228	12393
200					

Table 7

Normal Yield Table Values

DOUGLAS-FIR MODEL

The program developed for ponderosa pine on a site index of 100 can be changed to match the yield table growth for different site indexes or the program can be altered to fit other species of trees. To illustrate this flexibility, the changes will be made to duplicate the yield table values for West Coast Douglas-fir on a site index of 200, a very good site for this species west of the Cascade Divide in Oregon, Washington, and British Columbia.

1. The annual diameter growth, D1=.282, given in statement 20, becomes D1=.455.
2. The annual height growth, D2=2.3, given in statement 40, becomes D2=4.2.
3. The correction value for the effect of height on height growth, D3=.0006, given in statement 65, becomes D3=.000645.
4. The correction value for the effect of basal area on diameter growth, D3=.0006, given in statement 65, becomes D3=.000645.
5. The height at which D5 is changed, given in statement 140 as 100, becomes 200.
6. The value of D5, given in statement 150 as .0152 becomes .0175.

7. The beginning mortality, D=.0304, given in statement 15, becomes D=.071.

8. The new mortality value, given in statement 193 as D=.0245, becomes D=.037.

9. The starting value for computing volume, V=.26, given in statement 120, becomes V=.89.

10. The rate at which V2 changes each year betwen age 60 and 135 years, given in statement 180 as V2=V2-.0000055, becomes V2=V2-.000021 and applies to the period from 90 to 170 years. The slowdown in form class improvement starts later than it does for ponderosa pine and the reduction is greater.

11. The decrease in mortality as the tree becomes older was given in statement 200 with a ratio of 56.5 over the age applied as a multiplier to the mortality value. This reduction starts in the year 60 and D6=56.5 in statement 16 becomes D6=24.

With these changes, the program produces the results shown in Table 9 when run with 2,300 trees per acre as the beginning stocking level. Table 9 provides the comparable values from McCardle (1949). (Tarif volumes replace the yield table volumes). The entry tables used were Tarif Access Tables for Pacific Northwest Species, Volume 2, Major Coastal Species (Meyer 1938). The final program is listed in Table 8.

```
10 DIM P(201,5)
13 PRINT"HOW MANY TREES PER ACRE DO YOU WISH TO PLANT";
14 INPUT P(1,3)
15 D=.071
16 D6=24
20 D1=.455
30 D4=.00228
40 D2=4.2
50 D5=.0165
65 D3=.000645
70 PRINT" AGE   DBH   HEIGHT   NO. OF TREES BASAL AREA TOTAL VOL."
80 N=10
90 V2=.0006
100 FOR I=1 TO 200
110 IF P(I,1)<2 THEN 131
115 P(I,4)=(P(I,1)/2)**2*3.142/144*P(I,3)
120 V=.89
130 V1=V1+P(I,1)/P(I,2)
131 IF I<8 THEN 140
132 P(8,2)=4.5
140 IF P(1,2)<200 THEN 151
150 D5=.0175
151 IF P(I,4)<130 THEN 155
152 D9=D1-D3*P(I,4)
153 GO TO 160
155 D9=D1
160 IF I<90 THEN 190
170 IF I>170 THEN 190
180 V2=V2-.000021
190 V=V+V2*I
192 IF I<21 THEN 197
193 D=.037
194 IF P(I,4)<80 THEN 197
195 D0=.000005
196 GO TO 198
197 D0=0
198 D8=D+D0*(P(I,4)-80)
199 IF I<60 THEN 204
200 D8=D8*D6/I
204 IF I<8 THEN 220
205 P(I+1,1)=P(I,1)+D9-D4*P(I,1)
210 P(I+1,2)=P(I,2)+D2-D5*P(I,2)
220 P(I+1,3)=P(I,3)-D8*P(I,3)
225 P(I,5)=V*V1*P(I,1)*P(I,3)
230 IF I<>N THEN 270
240 PRINT I;TAB(6);INT(P(I,1)*100+.5)/100;TAB(13);INT(P(I,2)+.5);
250 PRINT TAB(25);INT(P(I,3)+.5);TAB(35);INT(P(I,4)+.5);
251 PRINT TAB(46); INT(P(I,5)+.5)
260 N=N+10
270 NEXT I
280 END
```

Table 8

Basic computer language for generating
Douglas fir yield tables for a site index of 200

```
HOW MANY TREES PER ACRE DO YOU WISH TO PLANT ?2300
 AGE    DBH    HEIGHT   NO. OF TREES BASAL AREA TOTAL VOL.
 10     0.91   13            1185         0            0
 20     5.39   50            568          90           2243
 30     9.2    81            375          173          6070
 40     12.26  103           255          210          8805
 50     15.12  130           174          217          10224
 60     17.92  149           119          203          10630
 70     20.6   165           103          238          13097
 80     23.04  179           91           264          15534
 90     25.28  191           82           285          17900
 100    27.35  200           74           302          19813
 110    29.26  207           68           317          21513
 120    31.05  212           63           329          22988
 130    32.72  217           58           339          24211
 140    34.29  220           54           343          25159
 150    35.78  224           51           355          25808
 160    37.19  226           43           362          26134
 170    38.53  229           45           367          26113
 180    39.81  230           43           371          27208
 190    41.03  232           41           375          28245
 200    42.2   233           39           379          29226
```

Table 9

Douglas fir normal yield table
for a site index of 200

LITERATURE CITED

Beck, Donald E. 1974. Predicting growth of individual trees in thinned stands of yellow poplar. Growth models for tree and stand simulation, pp. 47-55. Skegshöyekolan, Royal College of Forestry, Stockholm, Sweden.

Burkhart, Harold E., and M.R. Strub. 1974. A model for simulation of planted loblelly pine stands. Growth models for tree and stand simulation, pp. 128-135. Skegshöyekolan, Royal College of Forestry, Stockholm, Sweden.

Dahms, W. G. 1960. Long-suppressed ponderosa pine seedling respond to release. Research note Pacific Northwest Forest and Range Experiment Station.

Harlow, W. M., and E. S. Harrar. 1972. Textbook of dendrology. McGraw-Hill, New York.

Lin, J.Y. 1974. Stand growth simulation models for Douglas-fir and western hemlock in the Northwestern United States. Growth models for tree and stand simulation, pp.102-118. Skegshöyekolan, Royal College of Forestry, Stockholm, Sweden.

Little, Gene R. 1972. Tree-volume tarif access tables for Pacific Northwest species. Washington Department of Natural Resources, Olympia, Washington. Volume 2, 2.

Little, Gene R. 1972. Tree-volume tarif access tables for Pacific Northwest species. Volume 1 iuterior species. State of Washington, Department of Natural Resources, Olympia, Washington. Volume 1, 2.

McCardle, R. E., W. H. Meyer, and D. Bruce. 1949. Yield of Douglas-fir in the Pacific Northwest. USDA Tech. Bull. 201.

Meyer, Walter H. 1938. Yield of even-aged stands of ponderosa pine. USDA Tech. Bull. 630.

Smith, David Martyn. 1962. The practice of silviculture. 7th Edition. John Wiley & Sons, Inc., New York. p. 315.

Turnbull, Kenneth J., G.R. Little, and G.E. Hoyer. 1963. Comprehensive tree-volume tarif tables. Second Edition. Washington Department of Natural Resources, Olympia, Washington.

Young, Harold E., and P. M. Carpenter. 1967. Weight, nutrient element and productivity studies of seedlings and saplings of eight tree species in natural ecosystems. Main Agric. Exp. Stn. Tech. Bull. 28, Univ. of Maine.